

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 27, 2002, 13:45:31 Search time 39 Seconds

(without alignments)  
1701.508 Million cell updates/sec

Title: US-09-926-084-7

Sequence: 1 MDLSLFVVSVAVLVGVSSSHV.....KGEYFFLPISALTATIAA 498

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A.Geneseq.101002.\*  
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2581	99.2	498	21	Geotrichum candidu
2	185	7.1	421	21	Neisseria meningit
3	177	6.8	421	21	Neisseria meningit
4	174	6.7	421	21	Neisseria meningit
5	137.5	5.3	439	22	Neisseria gonorrhoe
6	130.5	5.0	423	22	Neisseria gonorrhoe
7	121.5	4.7	1093	23	Neisseria gonorrhoe
8	121	4.7	508	22	Neisseria gonorrhoe
9	118	4.5	2383	21	Neisseria gonorrhoe
10	114.5	4.4	567	22	Neisseria gonorrhoe

11	112.5	4.3	1837	21	Neisseria gonorrhoe
12	112.5	4.3	1837	23	Neisseria gonorrhoe
13	112	4.3	197	22	Neisseria gonorrhoe
14	111.5	4.3	523	22	Neisseria gonorrhoe
15	111.5	4.3	1194	22	Neisseria gonorrhoe
16	110	4.2	21	21	Neisseria gonorrhoe
17	110	4.2	1909	22	Neisseria gonorrhoe
18	109.5	4.2	421	23	Neisseria gonorrhoe
19	109.5	4.2	4924	22	Neisseria gonorrhoe
20	109.5	4.2	4928	20	Neisseria gonorrhoe
21	109	4.2	1600	22	Neisseria gonorrhoe
22	107	4.1	21	21	Neisseria gonorrhoe
23	107	4.1	940	23	Neisseria gonorrhoe
24	106.5	4.1	615	22	Neisseria gonorrhoe
25	105.5	4.1	729	22	Neisseria gonorrhoe
26	105	4.0	753	22	Neisseria gonorrhoe
27	105	4.0	2469	22	Neisseria gonorrhoe
28	105	4.0	3386	22	Neisseria gonorrhoe
29	104.5	4.0	1684	12	Neisseria gonorrhoe
30	104	3.9	603	14	Neisseria gonorrhoe
31	102.5	3.9	902	21	Neisseria gonorrhoe
32	102.5	3.9	1003	8	Neisseria gonorrhoe
33	102.5	3.9	1712	22	Neisseria gonorrhoe
34	102	3.9	513	9	Neisseria gonorrhoe
35	102	3.9	2089	17	Neisseria gonorrhoe
36	101.5	3.9	1003	7	Neisseria gonorrhoe
37	101	3.9	398	22	Neisseria gonorrhoe
38	101	3.9	409	22	Neisseria gonorrhoe
39	100	3.8	664	22	Neisseria gonorrhoe
40	100	3.8	664	22	Neisseria gonorrhoe
41	100	3.8	708	21	Neisseria gonorrhoe
42	100	3.8	905	22	Neisseria gonorrhoe
43	100	3.8	1022	9	Neisseria gonorrhoe
44	99	3.8	1389	21	Neisseria gonorrhoe
45	98.5	3.8	712	19	Neisseria gonorrhoe

## ALIGNMENTS

RESULT 1	AA1980.DAT	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
ID	AA1980.DAT	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
XX	AA1980.DAT	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
AC	AA1980.DAT	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
XX	AA1980.DAT	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
DT	09-JAN-2001 (first entry)	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
XX	09-JAN-2001 (first entry)	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
DE	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
XX	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
KW	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
XX	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
OS	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
XX	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
PN	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
XX	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
PD	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
XX	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
PP	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
PR	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
XX	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
PA	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
XX	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
PI	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
XX	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
DR	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
XX	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.
PT	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.	Geotrichum candidum Dec 1 (FERM BP-7033) DYP protein SEQ ID NO:7.



QY 471 TVPLFVLPKGEYFF-LPSI 489  
Db 393 ----YISPGGIFVLPGV 408

## RESULT 3

AA75545  
ID AA75545 standard; Protein; 421 AA.

XX AA75545;

XX 21-MAR-2000 (first entry)

XX Neisseria meningitidis ORF 748 protein sequence SEQ ID NO:2564.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;

KM antibacterial; gene therapy.

XX Neisseria meningitidis.

XX WO957280-A2.

XX 11-NOV-1999.

XX 30-APR-1999; 99WO-US09346.

XX 01-MAY-1998; 98US-0083758.

XX 31-JUL-1998; 98US-0094869.

XX 02-SEP-1998; 98US-0098994.

XX 02-SEP-1998; 98US-0098994.

XX 09-OCT-1998; 98US-0103749.

XX 09-OCT-1998; 98US-0103749.

XX 09-OCT-1998; 98US-0103749.

XX 25-FEB-1999; 99US-0121528.

XX (CHIR) CHIRON CORP.

XX (GENO-) INST GENOMIC RES.

XX Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;

XX Petersen J, Pizsa M, Rappunli R, Ratti G, Scalato E, Scarselli M;

XX Tettelin H, Venter JC;

XX WPI; 2000-062150/05.

XX N-PSDB; AA254307.

XX Novel Neisserial polypeptides predicted to be useful antigens for

XX vaccines and diagnostics

XX Claim 2, Page 1218, 1453pp; English.

XX AA253015 to AA254536, AA254577 to AA254615, and AA774253 to AA75941

XX represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides

XX and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent

XX PCR primers used in the exemplification of the present invention. The

XX polypeptides, the polynucleotides, antibodies and compositions of

XX the invention can be used as vaccines, as diagnostic reagents, and as

XX immunogenic compositions. The polypeptides can be used in the

XX manufacture of medicaments for treating or preventing infection due to

XX Neisserial bacteria (e.g. meningitis and septicemia), to detect the

XX presence of Neisseria bacteria, or to raise antibodies. They may also

XX be used to screen for agonists or antagonists, which may themselves

XX have use as antibacterial agents. The polynucleotides of the invention

XX may also be used in gene therapy protocols.

XX Sequence 421 AA;

XX Query Match 6.8%; Score 177; DB 21; Length 421;

XX Best Local Similarity 23.4%; Pred. No. 3,7e-07;

XX Matches 89; Conservative 41; Mismatches 124; Indels 126; Gaps 18;

XX 135 LGTTDLGNAQFPDQFA-----DAAN-----LGDSDLQVAVPFTGTTT 174

Db 130 LTVTVGVSSSLF-DGRFPLKDKKPIHLOEMKRDPSNDKLSQKWCDSLSIQICAFPTPTC 188  
QY 175 GVPLIGSDDDFLDOFTDDISSYFCSSITO--VOALSGARPPDQAGHEHFGFLDGISOP 232  
Db 189 AA-----LRDIIKHTVOTAVIRMSIDGMQPKSEPGAMARNLLGFRDGTGNP 235  
QY 233 SVTGMETTVFPGQAVVPFGIILITGRDDGTGRPSMALDGSFMAFRFQOKVEFNAYTLA 292  
Db 236 KVSDEKTA-----DEVLMTGVANSLDDEPMKANGSYQAVRLIRHFVFWERTPL- 285  
QY 293 NAIPIANSAGNLTOEGAEFLCARFMFGKSGAPIDLAFTADDPALGADPRNNPDYSOT 352  
Db 286 -----DEQDTIFGRKRY--SGAPMDCKKXADQDFAKDPE----- 318  
QY 353 LIDETRCPEGAHVRRKTNPRDLAGP--VDFPHARSSIPYGPETSDAEIASGVTAQDRGL 410  
Db 319 -GDIT--PDSHRIILANRD---PEFLKXGRLEFRAYSY--\$GLASSG--QADVGL 365  
QY 411 LFEVQSIITGNGFRFOQINMANANPPPSKPTPGLEPIIGOTTPTTGVGLDPLNOMETE 470  
Db 366 VFCVYQANLADGFIFVQ-NLNG-----EPLEB--- 392  
QY 471 TVPLFVLPKGEYFF-LPSI 489  
Db 393 ----YISPGGIFVLPGV 408

## RESULT 4

AA75544  
ID AA75544 standard; Protein; 421 AA.

XX AA75544;

XX 21-MAR-2000 (first entry)

XX Neisseria gonorrhoeae ORF 748 protein sequence SEQ ID NO:2562.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;

KM antibacterial; gene therapy.

XX Neisseria gonorrhoeae.

XX WO957280-A2.

XX 11-NOV-1999.

XX 30-APR-1999; 99WO-US09346.

XX 01-MAY-1998; 98US-0083758.

XX 31-JUL-1998; 98US-0094869.

XX 02-SEP-1998; 98US-0098994.

XX 02-SEP-1998; 98US-0098994.

XX 09-OCT-1998; 98US-0103749.

XX 09-OCT-1998; 98US-0103749.

XX 25-FEB-1999; 99US-0121528.

XX (CHIR) CHIRON CORP.

XX (GENO-) INST GENOMIC RES.

XX Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;

XX Petersen J, Pizsa M, Rappunli R, Ratti G, Scalato E, Scarselli M;

XX Tettelin H, Venter JC;

XX WPI; 2000-062150/05.

XX N-PSDB; AA254306.

XX Novel Neisserial polypeptides predicted to be useful antigens for

XX vaccines and diagnostics

XX Claim 2, Page 1217-1218, 1453pp; English.

```
XX AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AA275941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA254673 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisseria bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 421 AA;
Query Match 6.7%; Score 174; DB 21; Length 421;
Best Local Similarity 23.0%; Pred. No. 6.9e-07;
Matches 87; Conservative 39; Mismatches 130; Indels 122; Gaps 15;
QY 135 LGITDGLGDAQFPQCFPA-----DAAN-----LGDLSQWVAPFTGTH 174
DB 130 LRVTVGVGSUUF-DGRFLGDKKTVHLOEMRDFPNKLOKWCDDGLSLQICAFPTCQ 188
QY 175 GVFLIGSQDDFLDQFTDDISSTFGSSITQ--VQALSGSGARPPDQAGHEHFGFLDISQ 232
DB 189 -----TALRDIKHQAOTAVIRWSIDGNQPKSEPCAMAARNLLGRDGTGNP 235
QY 233 SVTGHETTVFGQAVVPGIILTCRDTGTRSPWALDGSFMAFRHQKVPFNAYTLA 292
DB 236 KVSDDPKTA-----DEVLWTCVAANSJDEPENAKNGSYQAVLIRFVFWORTPL- 285
QY 293 NAIIPANSAGNLTOQEGAEFLGARFGRKWSGAPIDLAPTADPALGADPQRNNNFDSGT 352
DB 286 -----QEQTDFGRKKV----SGAPMDGKKEADQDPFAKDPE----- 318
QY 353 LTDETRCFGAHVKRTNPRQDLGQPVDTFHMRSSIPYGPETSDAELASGVTAQDRGLLF 412
DB 319 -GDIT--PKDSHMLANPRDPEFLKXCHLFRFAYSYSRGPASSG-----QLDVGLVP 367
QY 413 VEYQIINGRFRFOIWMANNANPPFSKPTPGIEPIGQTPRTVGGDLPLNQNETPTV 472
DB 368 VCYOANLADGFIVQ--NLLNG-----EPLLE----- 392
QY 473 PLFVLPKGEYFF-LPSI 489
DB 393 --YISPFGGGTFVLPGV 408
```

```
RESULT 5
AAU36119
ID AAU36119 standard; Protein; 439 AA.
XX
AC AAU36119;
XX
DT 14-FEB-2002 (first entry)
XX
DE Klebsiella pneumoniae cellular proliferation protein #107.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
OS Klebsiella pneumoniae.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
```

```
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX (ELIT-) ELITRA PHARM INC.
XX
XX Hasebeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
DR WPI; 2001-611495/70.
DR N-PSDB; AAS53978.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
PT
XX
XX Example 3; Seq ID No 11712; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 439 AA;
```

```
Query Match 5.3%; Score 137.5; DB 22; Length 439;
Best Local Similarity 21.2%; Pred. No. 0.0014;
Matches 108; Conservative 47; Mismatches 160; Indels 195; Gaps 22;
```

```
QY 4 SLFVSVAVLVGSSSHVNAKLGNRQTRTPLLNTNPPGOAP-----LPTLTQHTTESGA 57
DB 87 SMLVAFDVLAAKADLE--RLPRLTQRIAFITQ--GGPAPDTPNRLPPM-----DSGI 138
QY 58 NDTILPLNNIQGDLVGMKKQKRFVFFQVNDATSEKTKALTYVPRTSAAILISDPSQ 117
DB 139 LCPMIAPDNLNLTITVSGHSLFDERF-----GLADKAPKKL----- 173
QY 118 OPLAFVNLGFNTGLQALGITDDLGAQDFDGFADANLGDLSQWVAPFTGTHCVF 177
DB 174 QPM-----TRFNSLDALCHGDLLOICANTODTVIHALR 210
QY 178 LTGSDQDDFL-----DQFTDDISSTFGSSITQVQALSGSARPPDQAGHEHFGFLDLSQ 231
DB 211 DVIKHTPDLLSVRWKREGFTSDSAARSKGKETPINLL-----GFKDGTAN 255
QY 232 PSVTGWETVPFGQAVVPGIILTGSDGTGTRPSWALDGSFMAFRHQKVPFNAYTL 291
DB 256 FA-----SHDSALMDKVVWVTADQD---EPTWTVGGSYQAARIIOFHV-EPWDRF- 301
QY 292 ANAIPANSAGNLTOQEGAEFLGARMEGRWK--SGAPIDLAPTADDPALGADPQRNNNFDS 350
DB 302 -----PLKEQ-----TIFGRDKHTGAPLGMKNEHDTF-----DYS 332
QY 351 DTLTDETRCFGAHVKRTNPRQDLGQPVDTFHMRSSIPYGPETSDA-----ELASGV 403
DB 333 KDPNGEV-IALDSHILANPRT-----PETOSSLMRGRYSYSLGV 372
```



QY 404 T---AODRGLFVEYQSIIGNGFRFOOINMANNANFPSPKRTPIGIEPIIGTTPRTVG 460  
 DB 373 TNAGOLDMGLFVCYQHLEKGFLLVCKRLNGEALVEYVPI----- 414  
 QY 461 LDPLNOMETFVPLFVIRKGEYFP-LPSI 489  
 DB 415 -----GGGYFVPLPGV 425

RESULT 6  
 AAU34528  
 ID AAU34528 standard; Protein; 423 AA.  
 XX  
 AC AAU34528;  
 XX  
 DT 14-FEB-2002 (first entry)  
 XX  
 DE E. coli cellular proliferation protein #109.  
 XX  
 KW Antisense; prokaryotic cellular proliferation protein;  
 KM antibiotic; antibacterial; drug design.  
 XX  
 OS Escherichia coli.  
 XX  
 PN WO200170955-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 21-MAR-2001; 2001WO-US09180.  
 XX  
 PR 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 PI Haselbeck R, Ohlsen KL, Zyekind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto R, Xu HH;  
 DR WPI; 2001-611495/70.  
 DR N-PSDB; AAS52387.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 XX Example 3; Seq ID No 10121; SLPD; English.

CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 CC Sequence 423 AA;

Query Match 5.0k; Score 130.5; DB 22; Length 423;  
 Best Local Similarity 22.3k; Pred. No. 0.0058;  
 Matches 81; Conservative 29; Mismatches 119; Indels 135; Gaps 14;

QY 145 QEPDGOFADANLGDGLDSQVAPPTGTTIHGVFLIGSDDDFL-----DOFTDISSTF 198  
 DB 162 REPNDSDLAALCHGCVLLQICANTQDTVHALRDIKHTPDLISYMKKEGFIIDHAARS 221  
 QY 199 GSSITVOVALSSARPPDQAGHEHFGFLDGISQPSVTG-----WETTPPGAAVPP 250  
 DB 222 KGETPIINLL-----GFKDGTAMPDSQNDKLMQKVVWTA----- 256  
 QY 251 GIILTRGDGTGTRPSMALDGSFMAFRHFOOKVPEFNAYTLNANIPANSAGNLTOEGAE 310  
 DB 257 -----DQGEPAWITGSSYOAVRLIOFRV-EMDRP-----PLKEQO--- 291  
 QY 311 FLGAMFGKWK-SCAPIDLAFTADPALGADPQRNNPFVSDTLDETCTCPGAAVHRTN 369  
 DB 292 ----TTFGRDKOTGAPLGMQHEHDVPPYASDPBCK-----VIALDSHILAN 334  
 QY 370 PRODLGGPYDTFHAMRSSIPYGPETSDAELASGVTAO---DRGLFVEYQSIIGNGFRQ 426  
 DB 335 PRT---AESSESLMRGYSY-----SLGVTNSGOLDMGLLFVCYQHDLEKGFLLV 382  
 QY 427 QINMANNANFPSPKRTPIGIEPIIGTTPRTVGGLDPLNOMETFVPLFVIRKGEYFP- 485  
 DB 383 QKRLNGEALVEYVPI-----GGGYFPA 405  
 QY 486 LPSI 489  
 DB 406 LPGV 409

RESULT 7  
 ABB54302  
 ID ABB54302 standard; Protein; 1093 AA.  
 XX  
 AC ABB54302;  
 XX  
 DT 16-MAY-2002 (first entry)  
 XX  
 DE Lactococcus lactis protein ykbc.  
 XX  
 KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.  
 XX  
 OS Lactococcus lactis IL1403.  
 XX  
 PN FR2607446-A1.  
 XX  
 PD 12-OCT-2001.  
 XX  
 PF 11-APR-2000; 2000FR-0004630.  
 XX  
 PR 11-APR-2000; 2000FR-0004630.  
 XX  
 PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.  
 PI Boletine A, Sorokine A, Renault P, Ehrlich SP;  
 PI WPI; 2002-043418/06.  
 DR  
 XX  
 PT New nucleotide sequence useful in the identification or Lactococcus  
 PT lactis and related species -  
 XX  
 PS Claim 6; SEQ ID No 1004; 2504bp; French.

CC The present invention is related to a Lactococcus lactis nucleotide  
 CC sequence (ABA50521) and related proteins (ABB53300-ABB55621). The  
 CC nucleic acid sequence is useful in the detection and/or amplification of  
 CC nucleic acid sequence, particularly to identify Lactococcus lactis or  
 CC related species. The proteins of the invention are useful for the  
 CC biosynthesis or biodegradation of a composition of interest. The  
 CC invention helps research in lactic bacteria, particularly useful in the



AAB15945  
 ID AAB15945 standard; Protein: 2383 AA.  
 XX  
 AC AAB15945;  
 XX  
 DT 05-OCT-2000 (first entry)  
 XX  
 DE E. coli proliferation associated protein sequence SEQ ID NO:302.  
 XX  
 KM Escherichia coli; E. coli; proliferation; inhibition; screening;  
 XX antimicrobial; bacterial growth; antisense therapy; antibacterial.  
 OS Escherichia coli.  
 XX  
 PN WO200044906-A2.  
 XX  
 PD 03-AUG-2000.  
 XX  
 PF 27-JAN-2000; 2000WO-US02200.  
 XX  
 PR 27-JAN-1999; 99US-0117405.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr GJ;  
 XX Yamamoto RT, Xu HH;  
 XX WPI; 2000-514822/46.  
 DR N-PSDB; AAA65950.  
 XX  
 PT Novel polynucleotides and polypeptides associated with microorganism  
 PT proliferation, used to identify inhibitors of bacterial growth and  
 PT proliferation, for use in antisense therapy -  
 PS Claim 11; Page 224-229; 316pp; English.  
 XX  
 CC AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide  
 CC sequences derived from Escherichia coli which inhibit E. coli  
 CC proliferation. AAA65890 to AAA66055 and AAB15886 to AAB16040 represent  
 CC nucleotide and protein sequences associated with E. coli proliferation.  
 CC AAA66056 and AAA66057 represent primers used for sequencing E. coli  
 CC proliferation inhibiting nucleotide inserts in an example from the  
 CC present invention. Methods from the present invention can be used to  
 CC identify a proliferation- required gene in a microorganism, by contacting  
 CC a microorganism with a proliferation-required gene activity inhibitory  
 CC nucleic acid identified in another organism, and determining if  
 CC inhibition occurs in the second microorganism. The nucleic acid sequences  
 CC identified as being required for bacterial growth and proliferation, can  
 CC be used for antisense therapy for killing bacteria.  
 CC  
 SQ Sequence 2383 AA;  
 Query Match 4.5%; Score 118; DB 21; Length 2383;  
 Best Local Similarity 19.5%; Pred. No. 1;  
 Matches 95; Conservative 54; Mismatches 160; Indels 178; Gaps 19;  
 QY 106 TSAALLISPSQOQPLAFVNLGFS-----NGL-----QALG----- 136  
 DB 1086 TLTATATVQKPSNHPVAGITVNTPOVYANFTLBNNGIATLNGAHVTLTKGKAGTN 1145  
 QY 137 -ITDDIGDAQFPDQ--FADANAIAGDLSQWVAPFTGTTINGVPLIGSDODFLDQFT 191  
 DB 1146 TVYATLTGNNTSDSQVPTVADKASQVVLQISKDEITGNQVDSATLTATVQKQF-DNEV 1204  
 QY 192 DDISFFGSSITQVQALSGSARPRQAGHEHFGFLDGISQSVYGMWTTVPQDAVPPG 251  
 DB 1205 NNLPTVFFSSASGLTITPCVSTN-----SGIAQVTLG-----VAFSEKTVLAS 1250  
 QY 252 IILTGSDGDTGRPSWALDGSFMAERHQVQKVEFNATYLANAIPANSAGNLTOQSGAEF 311  
 DB 1251 IANNGAS-----DNKTVHFICDTAAKI----- 1273  
 QY 312 LGARMGRKMSGAPIDLAFTADDPALGADPQRNNPFQVSDTLTDETSCPF----- 361

DB 1274 -----TELAPVPDSIIAGT-PONSSGVITATVDNNGFVKQVTVNFTSN 1318  
 QY 362 GAHVRRKNPRQ---DLGGPDTFPHAMRSSIPYG--PETSDELASG----- 402  
 DB 1319 AATAEWTNGGAVTNEQSKATVYTNTRSSIESGAPDPTVEASLENGSSTLTSINVAN 1378  
 QY 403 -----VTNODRGLFVEYQSIIGNGRFOQIMANNANPPFSKPTIPGIE 447  
 DB 1379 ASTAHLTLQALFDIVSAGETTSIXIEVKDNYNGNVPOQEVTLISVP-----SEGVTPSN 1434  
 QY 448 PIIQOTTPRTVGGLDPLNQNFTVPLFVIPKGEY-----FLPFSISA 491  
 DB 1435 AI--YTT-----NHDGNFYAS-FYATKAGVYQVLTATLENGSDPMOQTVYVPNVAN 1481  
 QY 492 LTATIAA 498  
 DB 1482 AETTLAA 1488  
 RESULT 10  
 ABE28940  
 ID ABE28940 standard; Protein: 567 AA.  
 XX  
 AC ABE28940;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #28931.  
 XX  
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS93127.  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT biologically for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PS Claim 20; SEQ ID NO 59299; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 567 AA;  
 Query Match 4.4%; Score 114.5; DB 22; Length 567;  
 Best Local Similarity 24.2%; Pred. No. 0.25;  
 Matches 55; Conservative 20; Mismatches 67; Indels 85; Gaps 9;  
 QY 265 PSWALDGSFMAFRHQKVFEPNAYTLAIPANSAGNLTOEGAEFLGARMFGRWK-SG 323  
 DB 83 PAWTIGGSQVAVLIQFRV-EFWDRT-----PLKEQO-----TIFGRDKQTG 123  
 QY 324 APIDLAPTADDPALGADPQRNNFDYSDTLTDETRCPFGAHVRKTNPRDLGGPVDTPHA 383  
 DB 124 APLGQHGHDPVYASDECK-----VIALDSHRLANPT---AESSESLM 167  
 QY 384 MRSSIPYGPETSDALASGVTAQ---DRGLLFVEYQSIIQNGFRFQOINWANNANFPESK 440  
 DB 168 LARGYSY-----SLGVNTSQGLDMGLLFVCYQHDLEKGFLLTVQKRLNGEALBEYVK 218  
 QY 441 PITPGIEPIIGQTTPRTVGGDLPLNQNETFTVPLFVIPKGGYFFLP 487  
 DB 219 PI-----GGGYFFAP 228

RESULT 11  
 ID AAB11726 standard; Protein; 1837 AA.  
 AC AAB11726;  
 XX  
 DT 28-OCT-2000 (first entry)  
 DE Cryptosporidium parvum Iowa isolate GP900.  
 KW GP900; Iowa isolate; glycoprotein; antibody; cryptosporidiosis;  
 KW competitive inhibition; attachment; invasion; ligand binding; sporozoite;  
 KW merozoite; diarrhoea; protozoacide.  
 OS Cryptosporidium parvum.  
 XX  
 PH Key Location/Qualifiers  
 FT Misc-difference 1785  
 FT /note= "Encoded by AAC in AAA61846"  
 XX  
 PN US6071518-A.  
 XX  
 PD 06-JUN-2000.  
 XX  
 PF 12-SEP-1997; 97US-0928361.  
 XX  
 PR 13-SEP-1996; 96US-0026062.  
 PR 01-JUN-1993; 93US-0071880.  
 PR 29-MAY-1992; 92US-0891301.  
 PR 03-APR-1995; 95US-0415751.  
 PR 14-AUG-1996; 96US-0700651.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Petersen C;  
 XX  
 DR WPI; 2000-422065/36.  
 DR N-PSDB; AAA61846, AAA61847.  
 XX  
 PT New GP900 protein fragments and fusion proteins of Cryptosporidium  
 PT parvum, useful for detecting the presence of the parasite, and  
 PT diagnosing or treating Cryptosporidium infections by competitive  
 PT inhibition of the function of GP900 -

XX  
 PS Claim 2; Column 51-60; 59pp; English.  
 XX  
 CC The invention relates to the GP900 glycoprotein of the protozoan  
 CC Cryptosporidium parvum, DNA encoding it, GP900 fragments, and fusion  
 CC proteins comprising GP900 fragments. The invention also relates to the  
 CC administration of GP900 or fragments thereof to a host to elicit anti-  
 CC GP900 antibody production, and to a method of cryptosporidiosis treatment  
 CC or prophylaxis comprising administration of anti-GP900 antibodies to an  
 CC individual. Cryptosporidium parvum GP900 and GP900 fragments are able to  
 CC competitively inhibit sporozoite or merozoite attachment or invasion, and  
 CC are also useful for the generation of anti-GP900 antibodies. The  
 CC antibodies also inhibit sporozoite or merozoite attachment/invasion and  
 CC additionally inhibit the binding of GP900 ligands to GP900. GP900  
 CC proteins, fragments and antibodies may therefore be used to treat or  
 CC prevent cryptosporidiosis. Infection with Cryptosporidium is a common  
 CC cause of diarrhoea in humans and causes life-threatening diarrhoea in  
 CC immunocompromised persons. Cryptosporidiosis can be contracted from  
 CC contaminated municipal water supplies (e.g., public swimming pools). It  
 CC is also a cause of disease in animals, resulting in financial losses in  
 CC agriculture. GP900 fragments, fusion proteins and antibodies may also be  
 CC used for the diagnosis of Cryptosporidium parvum infections, and for the  
 CC detection of the parasite in the environment. The present sequence  
 CC represents the GP900 protein of the Iowa isolate of Cryptosporidium  
 CC parvum.  
 XX  
 SQ Sequence 1837 AA;  
 Query Match 4.3%; Score 112.5; DB 21; Length 1837;  
 Best Local Similarity 18.5%; Pred. No. 2.2; 169; Indels 245; Gaps 30;  
 Matches 108; Conservative 64; Mismatches 64; Indels 245; Gaps 30;  
 QY 29 QTRTPLLTFPPGOAPL-----PTLTQHTTESGANDTIL 62  
 DB 736 OKPTTTTTTKVPGKPIATTTTTLKPIVTTTTTKATTTTTTPTTTTKRDEMTTTT 795  
 QY 63 PLNNIOGDI---LVGMEKKOKERF--VFQVNDATSEFKALKTYVPQRITSAALISPSQ 117  
 DB 796 PLPDI-GDIEITPIPIEMJDKYTRMYDYSNG-----LLLOSND 834  
 QY 118 QPLAFVNLGFSNTGLQALGITDDLGDAPFDQGFADAANL----- 157  
 DB 835 EPI-----PGSQA-----GQIADTSNLFVPVQTHKSTGLPIDPMVGLP 871  
 QY 158 -GDDLISQWVAPFTGTTJGV-----FLIGSDQ----- 183  
 DB 872 FDPKSGNLVHPYTNQWSLSVSYLAAKNLTVDTETVGLPIDTLTGYPDPVSLIPNP 931  
 QY 184 --DDFLDQFTDDI--SSTFGSSITQVOA----LSGSARPPDQAGHEHFLDGI SOPSVTG 236  
 DB 932 ETGELFDPISDEIMNGTITAGIVSGISASESLLSQSAPIDPATNMVVGEGGLLNPA-TG 990  
 QY 237 WETTVPFGQ-----AVVPPGIILTRGDGTGTRPSWALDGSFWAFRHQ 280  
 DB 991 ---VMIFGSLCPSEQTFPSEIEDGGIIPPEVAANADKFKLSIPSPVE---SIEKD 1043  
 QY 281 QKVPEFN--AYTLANA-----IPANSAGNL-----TOEGAEFLGARMFGRWK 321  
 DB 1044 QKIDSISELMYDIESGRLIQVSKRPIPGSIAGDLNPKMTPTQDSV----- 1091  
 QY 322 SGAPIDLAPTAD--DPALG--ADPORNNDYSDTLTDETRCPFGAH-----VRKTN 369  
 DB 1092 TKCPID--PTTGLFPNPPTCHLINPTNNNTMSSFA-----GATKYAVSNGIKTDN 1140  
 QY 370 PRQDLGGPVDTFHAMR-----SSIPYGPETSDA-----ELASGVTAQDRGLLFVEY 415  
 DB 1141 V---YGLPVDIEITGLPKDPVSDIPFNSTTGELVDPSTGKPINNYTAGIVSGKRGUPTIED 1197  
 QY 416 OSIIGNFRFOQINWANNANFPESKPIITPGIEPIIGQTTPRTVGG 460  
 DB 1198 E-----NGNLF-----PSTKLPIDGNQVLNPNETNIVSG 1228

RESULT 12  
ABJ04044 standard; Protein: 1837 AA.  
ID ABJ04044 standard; Protein: 1837 AA.  
XX  
AC ABJ04044;  
XX  
DT 27-SEP-2002 (first entry)  
XX  
DE C parvum GP900 protein fragment SEQ ID NO: 5.  
XX  
KM Cryptosporidium detection; GP900; P68; cryptopain; cryptosporidiosis.  
XX  
OS Cryptosporidium parvum.  
XX  
PN WO200194631-A1.  
XX  
PD 13-DEC-2001.  
XX  
PF 14-MAY-2001; 2001WO-US15624.  
XX  
PR 06-JUN-2000; 2000US-0588995.  
XX  
PA (REGC ) UNIV CALIFORNIA.  
XX  
PI Petersen C, Barnes DA, Nelson RG, Gut J;  
XX  
DR WPI: 2002-566447/60.  
XX  
PT Detecting Cryptosporidium in biological and environmental samples and  
XX  
PT diagnosis of cryptosporidiosis involves, contacting the sample with  
XX  
PT Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or RNA  
XX  
PS Disclosure: Page 104-110; 157pp; English.  
XX  
XX  
XX The present invention relates to a method of detecting Cryptosporidium in  
XX  
XX biological and environmental samples, and of diagnosing  
XX  
XX cryptosporidiosis. This involves obtaining a sample and contacting it  
XX  
XX with Cryptosporidium GP900, P68 or cryptopain antigen, antibody, DNA or  
XX  
XX RNA, or its variant, mutant or fragment. The method is also useful for  
XX  
XX detecting and identifying individual Cryptosporidium isolates based on  
XX  
XX the genetic characteristics, and for diagnosis of prior or concurrent  
XX  
XX Cryptosporidium infection. The present sequence is a C. parvum protein  
XX  
XX sequence used in the exemplification of the invention.  
XX  
SQ Sequence 1837 AA;  
XX  
Query Match 4.3%; Score 112.5; DB 23; Length 1837;  
Best Local Similarity 18.5%; Pred. No. 2.2;  
Matches 108; Conservative 64; Mismatches 168; Indels 245; Gaps 30;  
XX  
QY 29 QRTTLPPLNFGQAFL-----PTLTQHTTSGANDTIL 62  
DB 736 QKFTTTTIVKVGKXPPIATTTTTLKPIVTTTTTATTTTTPVTTTTTKRDEMTTTTT 795  
QY 63 PLNNIOGDI---LVGKKOKERF--VFQVNDATSFYAKTKVYVQRTSAALLIDPSQ 117  
DB 796 PLPDI-GDIEITPIPIEKMDLKYTMIDVNSG-----LLIDNSD 834  
QY 118 QPIAFVNLGFSNTGLQAGITDLDAGQFPGQPADAAAL----- 157  
DB 835 EPI-----PGSQ-----GQIADTSNLFVQTHKSTGLPIDWVGLP 871  
QY 158 -GDLSQWVAPFTGTIGV-----FLIGSDO----- 183  
DB 872 FDPKSNLVHPTNTQMSLSVYLAANKLVVDERTYGPIDLTIGYPLDPSLFPNP 931  
QY 184 --DPLDQETDDI-SSTFSSITQQA---LGGANRPFQAGHEHFGPLDGSQSVTG 236  
DB 932 ETGELPDPISDEIMNGTLAGIVSGISABESLSQSAFIDPATMVVGERGGLNPA-TG 990  
QY 237 METTVPGQ-----AVVPGIILTGRDGTGTSPSWALDGSFMAFRHQ 280

DB 991 ---VMIPSGLPSEQTPSPETEDGIIPEVAAANADKFKLSIPSPVE----SIPKED 1043  
QY 281 QKVPFEN--AYTLANA-----IPANSAGNL-----TQDGAFFLQARFGKWK 321  
DB 1044 QKIDISISELMTDIESGRLLIGVSKRPISGIDAGDLNPIKTPYDTDSV----- 1091  
QY 322 SGAPIDLAPAD--DPAIG--ADPORNNDVYSDDLTDETRCPPGAA-----VRKTN 369  
DB 1092 TKGPID--PTTGLPFPNPTGLINPTNNNTMDSSFA-----GAYVYAVSNGIKTDN 1140  
QY 370 PRQDLGGPVDTFHAMR-----SIPYGPETSDA-----ELASGVTAQDRGLLPVEY 415  
DB 1141 V---YGLPVDEITGLPKDPVSDIPENSTGSELVDPSTGKRPINNYTAGIVSGRGR,LPLED 1197  
QY 416 OSIIGNGFRFQOIWMANNANFPFSKPTTPGIEPIIGQTPRTYVG 460  
DB 1198 E---NGNLF-----PSTKLPIDNNQLVNPEPNTSVSG 1228  
RESULT 13  
ABG28941  
ID ABG28941 standard; Protein: 197 AA.  
XX  
AC ABG28941;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #28932.  
XX  
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;  
XX  
XX food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
XX  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
XX  
XX 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI: 2001-639362/73.  
XX  
XX N-PSDB; AAS93128.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
XX  
PT diagnostics, forensics, gene mapping, identification of mutations  
XX  
PT responsible for genetic disorders or other traits and to assess  
XX  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID NO 59300; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
XX  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX  
XX and gene mapping, and in recombinant production of (II). The  
XX  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX  
XX to restore normal activity of (II) or to treat disease states involving  
XX  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX  
XX quantitating a polypeptide in tissues, as molecular weight markers and as  
XX  
XX a food supplement. (II) and its binding partners are useful in medical  
XX  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX  
XX disorders involving aberrant protein expression or biological activity.  
XX  
XX The polypeptide and polynucleotide sequences have applications in  
XX  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX  
XX and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 197 AA;  
Query Match 4.3%; Score 112; DB 22; Length 197;  
Best Local Similarity 24.3%; Pred. No. 0.085;  
Matches 56; Conservative 21; Mismatches 67; Indels 86; Gaps 10;  
QY 265 PSWALDGSFMAFRHFOQVPEFNAYTLANAI PANSAAGNLTOEGAEFLGARMFGKWK-SG 323  
DB 35 PAWTIGGSYQAVRLIQFRV-EFWDR-----PLKEQ-----TFGRDKQTG 75  
QY 324 APIDLAPTADOPALGADPQRNNFDYSDTLTDETRCPFGAHVRKTNPRQDLGGPDVDTFHA 383  
DB 76 AFLGQHEDHVDYASDPEG-----VIALDSHRLANPRT---AESESLM 119  
QY 384 MRSSIPYGPETSDAELASCVTAQ---DRGLLVEYQSIIGNGRFOQINWANNANFPESK 440  
DB 120 LRGRYSY-----SUGVTNSQGLMGULFVCYQHDLEKGLTVQKRLNGEALERYVK 170  
QY 441 PITPGIEPIIGQTPRTVGGDLPLNQNETFVLPVIPKGEYFF-LPSI 489  
DB 171 PI-----GGGYFFALPGV 183  
RESULT 14  
AAB95114  
ID AAB95114 standard; Protein; 523 AA.  
AC AAB95114;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human protein sequence SEQ ID NO:17099.  
XX  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX  
XX Homo sapiens.  
XX  
XX EP1074617-A2.  
XX  
XX 07-FEB-2001.  
XX  
XX 28-JUL-2000; 2000EP-0116126.  
XX  
XX 29-JUL-1999; 99JP-0248036.  
XX  
XX 27-AUG-1999; 99JP-0300253.  
XX  
XX 11-JAN-2000; 2000JP-0118776.  
XX  
XX 02-MAY-2000; 2000JP-0183767.  
XX  
XX 09-JUN-2000; 2000JP-0241899.  
XX  
XX (HELI-) HELIX RES INST.  
XX  
XX Ota T, Isogai T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
XX full-length cDNAs defined in the specification, and for the detection  
XX and/or diagnosis of the abnormality of the proteins encoded by the  
XX full-length cDNAs -  
XX  
XX Claim 8; SEQ ID 17099; 2537pp + CD ROM; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602  
XX full-length cDNAs defined in the specification. Where a primer set  
XX comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
XX to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
CC AAH95993 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 523 AA;  
Query Match 4.3%; Score 111.5; DB 22; Length 523;  
Best Local Similarity 23.3%; Pred. No. 0.41;  
Matches 84; Conservative 38; Mismatches 148; Indels 91; Gaps 17;  
QY 91 TSFKTALKTVVPOKITSA-----ILISDP--SOOLAFVNLGFSNTGLQALGTD 139  
DB 171 TFSQSLLPTAVPTATSSNAADPSGFGSLTATAPATSSQP-----TLTFSNTSTPTFNI-- 224  
QY 140 DLGDAQPPDQGFADANLGDLLSQWVAPPTGTHGVFLIGSDQDFLDQTDTSSTFG 199  
DB 225 -----PFGSSAKSP-----LPSYPGANPQPTF--GAAGQPPCAAKPALAPSPG 266  
QY 200 SSITQVOALSGSARP-----PDAGHEHFGF-----LDGI 229  
DB 267 SSFTFGNSAAPAAAPTAPESMIKVVAHVPTPIHPFGGTHSAFGLKATASAFGAPAS 326  
QY 230 SQPSVTGWETTVFPGQAVPFGIILTRGDGTGTRPSWALDGSFMAFRHFOQVPEFNAY 289  
DB 327 SQPAP-GGSTAVFFG-AATSSGFGATTQTASSGSSS-----VFGSTTP--SPF 371  
QY 290 TL-ANAI PANSAG---NLTOEGAEFLGARMFGKWKSGAPIDLAPTA---DDPALGADPQ 342  
DB 372 IFGSAAPAGSGSGFINVATPGSSTTTGAFSGAGSGSSTATSTPFAGLGONALGTTGQ 431  
QY 343 RNNRPDYSDTLTDETRCPFGAHVRKTNPRQDLGGPDVDTFHAMRSSIPYGPETSDAELASG 402  
DB 432 -STPPAFNVSTTESKPVFG--TATPTFGLNTPAGVGTGSGSLSFSGASSAPAQPFVG 487  
QY 403 V 403  
DB 488 V 488  
RESULT 15  
ABG23389  
ID ABG23389 standard; Protein; 1194 AA.  
XX  
XX AC ABG23389;  
XX  
XX 18-FEB-2002 (first entry)  
XX  
XX Novel human diagnostic protein #23380.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
XX  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PA (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

PI

XX MPI: 2001-639362/73.

DR N-PSDB; AAS87576.

XX

XX

XX

XX

XX

PS Claim 20; SEQ ID No 53748; 103bp; English.

XX

XX

XX

CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (II) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG030377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

XX

SQ Sequence 1194 AA;

Query Match 4.3%; Score 111.5; DB 22; Length 1194;

Best Local Similarity 23.3%; Pred. No. 1.4;

Matches 84; Conservative 38; Mismatches 148; Indels 91; Gaps 17;

QY 91 TSPFKALKTYVPORITSA-----ILISDP--SQOPLAFVNLGFSNTGLQALGITD 139

DB 842 TTTSQSLHTAVPFAATSSADPSGSGTLATSAPATSSQP---TLTFSNTSTPTPTNI-- 895

QY 140 DLGDAQPPDQGFADANLGLDLSQWVAPFTGTHGVFLIGSDDDFLDQFTDDISSTFG 199

DB 896 -----PFGSSSKSP-----LPSYPGANPOPAF--GAAGQPPGAAXKPALAPSPG 937

QY 200 SSTVOVALSGSARP-----FDOAGHEHFGF-----LDGI 229

DB 938 SSFTFGNSAAPAAP 997

QY 230 SQPSVTGMEITVPPGQAVVPPCIILTGRDGTGRPSWALDGSFMAFRHFQKYPENAY 289

DB 998 SQPAF--GSGTAFFFG--AATSSGFGATTQTASSGSSS-----VFGSTTP--SPF 1042

QY 290 TL-ANAPANSAG--NLTOGEGAEFLGARFGRWKSAPIDLAPTA---DDPALGADPO 342

DB 1043 TFGGSAAPAGSGSGFGINVAIPGSSSTTTGAFSGAGQSGSTATSTPFAIGLGNALGTTGQ 1102

QY 343 RNNNFDYSDTLDETCTCPFGAHVKTNPQDLCGPVDTFHAMRSSIPYGPETSDAELASG 402

DB 1103 -STPFAFNVSSTTESKPVGG--TATPTFGINTPAPGVGTGSSLSFGASAPAGQFVG 1158

QY 403 V 403

DB 1159 V 1159

Search completed: November 27, 2002, 13:47:07  
JOB time : 44 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 27, 2002, 13:45:31 ; Search time 22 Seconds  
(without alignments)  
2176.134 Million cell updates/sec

Title: US-09-926-084-7

Sequence: 1 MDLFLPVSVAVLVGSSSHV.....KGGYFFLPSISALMTATIA 498

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	412.5	15.9	460	2	D75610 probable peroxidase
2	386.5	14.9	469	2	AC2004 hypothetical prote
3	185	7.1	421	2	B82023 probable periplasm
4	177	6.8	421	2	F81244 conserved hypothet
5	143.5	5.5	434	2	AE0226 conserved hypothet
6	133.5	5.1	423	2	A89787 hypothetical prote
7	133.5	5.1	423	2	A85647 hypothetical prote
8	130.5	5.0	423	2	A64844 ycdb protein precu
9	127.5	4.9	1067	2	D75625 conserved hypothet
10	126.5	4.9	299	2	AH0367 hypothetical prote
11	121.5	4.7	1093	2	B86748 hypothetical prote
12	119.5	4.6	420	2	T50585 probable membrane
13	118.5	4.6	2468	2	A83412 hypothetical prote
14	118	4.5	2383	2	D64962 probable membrane
15	116	4.5	2660	2	B85822 probable invasin Z
16	115	4.4	3972	2	S75251 hypothetical prote
17	112.5	4.3	1248	2	B96827 hypothetical prote
18	112.5	4.3	1832	2	T31113 mucin-like glycopr
19	111.5	4.3	2712	2	T05113 hypothetical prote
20	110	4.2	4199	2	S76412 hypothetical prote
21	109.5	4.2	421	2	AC1481 conserved hypothet
22	109.5	4.2	421	2	AH1120 B. subtilis twbn p
23	109.5	4.2	1742	2	S76110 hypothetical prote
24	108.5	4.2	3029	2	S76109 exo-alpha-glucidas
25	108	4.2	647	2	A45244 proline-rich prote
26	108	4.2	778	2	T17679 hypothetical prote
27	108	4.2	1345	2	H90975 probable ppe prote
28	107.5	4.1	2523	2	F70846 intermalin protein
29	107	4.1	940	2	AD1374

30	107	4.1	3562	2	A47171 chondroitin sulfat
31	106	4.1	743	2	T34853 probable fusidic a
32	105.5	4.1	700	2	H85732 hypothetical prote
33	105.5	4.1	859	2	AE2217 hypothetical prote
34	105.5	4.1	2232	2	T34434 hypothetical prote
35	105	4.0	753	2	B85782 catalase, hydropex
36	105	4.0	405	2	F90933 cell surface-asso
37	105	4.0	835	2	JC6140 fat protein - syn
38	105	4.0	1965	2	S75200 hypothetical prote
39	104.5	4.0	409	2	EB3179 NUP42 protein - ye
40	104.5	4.0	430	2	S52700 amylase A-180 - al
41	104.5	4.0	1684	2	S10789 basic serine prote
42	104	4.0	603	2	S27055 dipeptide ABC tran
43	104	4.0	658	2	H84385 catalase (EC 1.11.
44	104	4.0	753	2	A39129 extracellular seri
45	104	4.0	2554	2	AB3528

## ALIGNMENTS

RESULT 1	
D75610	probable peroxidase - Deinococcus radiodurans (strain R1)
C/Species: Deinococcus radiodurans	
C/Date: 03-Dec-1999	#sequence_revision 03-Dec-1999 #ext_change 02-Jun-2000
C/Accession: D75610	
R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; J	
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.	
Science 286, 1571-1577, 1999	
A/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.	
A/Reference number: A75250; MUID:20036896; PMID:10567266	
A/Accession: D75610	
A/Status: preliminary	
A/Molecule type: DNA	
A/Residues: 1-460 <WHI>	
A/Cross-references: GB:AE001862; GB:AE001825; NID:G6460468; PIDN:AAF12218.1; PID:G646051.	
A/Experimental source: strain R1	
C/Genetics:	
A/Map position: 2	
C/Superfamily: Deinococcus radiodurans probable peroxidase DRA0145	
Query Match	15.9%; Score 412.5; DB 2; Length 460;
Best Local Similarity	29.3%; Pred. No. 1.1e-23;
Matches 148; Conservative 68; Mismatches 190; Indels 99; Gaps 25;	
QY 34 PLITNPGQAPPTLTQHTSSGANDTI-LPNNIQDILVKKKKKER-FVFQVNDAT 91	
DB 2 PEWTLF--KKRLRELVNH-----NDKIDLDLDIOATVL-----REPREPYGTHANV 46	
QY 92 SFKTA-----IKTVPPRITSAAIILSDPSQPLAFVNFVGSNTGLQALITDDIGDAQ 145	
DB 47 RPTTAGGRELKRLP-HIASA-----EKWMDVAVMTAAAIIVEGKKLGVQSDLS- 100	
QY 146 FRP-----GQFADANLGD-----DLSQVAPFTGTTINGVLISDDDD-----FLDPTTD 193	
DB 101 FPFSPFVGNAGRAEHLFDVGENDPRGWKPRGTCGVLAULTIYFAEENMKQALVIAEHE 160	
QY 194 ISSTFSSITVOALSGSARPPDQAGHEHFGFLDGIQSPSVTGMETTVFPGQA-VVPPGI 252	
DB 161 LEATKQ--VTLLMRDFGAQP---DSRNSLGKYMISNPALIESGKIKPFQGPALKPPE 215	
QY 253 ILTGRGDGT-----RPS-WALDGSFMAFHPQCKPPEFNATTLNAAIPANSAGULTQO 306	
DB 216 FVLGYGEGAGVLPKPKPEVLGKNGTFVALRKRYTNAGSNRYLKEKNA--EYVGS- 269	
QY 307 EGAEFLGAMPGRWKSGAPIDLAADPALGADPQRNNPDDVSDTLTDETRCPGGAHYR 366	
DB 270 --AELLAAKLVGRWSGAPLTLAPKEDDPBLGHDPNRNNDFYKND--PEGLVPLCSHIR 326	
QY 367 KTNPRQ---DLGGPVDTHAMKSSIPYGPBTSDELAASGVTADRCGLLTFVYGSIIIGNGF 423	

Db 327 RMNPRDTELELLTVDNIHRIIRATAYGPAYDPKADSLAEDKVERGLYFIFISAKAMDTT 386  
Qy 424 RFOQINWANNANFFPSKPIPTGIEPIIGQTPRTVGGLOPL-----NONETFTVP----- 473  
Db 387 EFLQEWINKANF-----IQOSER-----DPFVGLQDEDLTFLPKPEVRQ 428  
Qy 474 -----LFVIPKGGEYFFLPSISAL 492  
Db 429 RLRGMDTFNVLRGGEYLFMPSLSLAL 453  
RESULT 2  
AC2004  
hypothetical protein alr1585 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp.  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AC2004  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguch  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AC2004  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-469 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA877951.1; PID:gl7135405; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr1585

Query Match 14.9%; Score 386.5; DB 2; Length 469;  
Best Local Similarity 29.9%; Pred. No. 1.1e-21;  
Matches 144; Conservative 57; Mismatches 185; Indels 95; Gaps 23;  
Qy 64 LNNITGDLVKKQKQKRFV--FQVNDATSFYKALKTVVQRTSA-----ALLISDP 115  
Db 28 LNDLQGNILKGRDHSVHLFQPKPEQVEVVKQWIQSQAQTVYTSAKKQADEAPKYRQK 87  
Qy 116 SQQLAFVNLGFSNTGLQALGIT--DDLGDQAPPDGQFAD--AANLGD--DLSQWVAPFTG 170  
Db 88 GVSQDVANFPLSRHGYEYLEIEPQIPGDKPFRMGKNEIRSSILGDPKIATWELGQ-- 146  
Qy 171 TTIHGVFLIGSDQDDFLDQFTDDISSTFGSSITQVQALSGSARPF---DOAGH--EHRGF 225  
Db 147 SEIHALVLIADDDIVDLQVQIT-----OKLQIAIEIVHREDGFIILRNQAGQIIIEHGF 202  
Qy 226 LDGISQPSVTGWTTFVFGQAVV-----PPGIILTRGDGDTGTRPSWAL 269  
Db 203 VDGVSQP-----LFMKRDVVRVNNCDPKDPKAPLDSILV--EDPNGTKDSY-- 251  
Qy 270 DGSFNAFRHFOOKVPEFNAYTLANIPANSAGNLTQEGAEFLGARMFGRWKSGAPIDLA 329  
Db 252 -GSYLIVYRKLQNKVAREDOUKLQKLNLTQENLA-----GALIVGRFADGTPVTL- 301  
Qy 330 PTADDPALGADPQRNNFDYSDTLTDTRCPFGAHVRKTNPRQDLGGPVDT----- 380  
Db 302 --SDIPTVAVTP--TNNFNVDGDLA--ATKCPFHSHTRKTNPRGDTARLLTTDGHFDEAPK 356  
Qy 381 ----PHAMRSIPYGPETSDAELASGVTAQDRGLLEVEVOSLIIGNCFQFQINWANNANF 436  
Db 357 EERGHIRITRAVSGENNPSPKPSVSG-----SGLLFLCQSNIEHQFNQMRWANPNQF 411  
Qy 437 PFSKPIITGPIIGQTT-----PRTVGGDLPLNQNETTFLVLPVKGGEYFFLPSISA 491  
Db 412 ---VQVNTGPDPLQPSGTOKWPKKWG--EP--ETEEYNFQNLWINMKKGGEYFFAPPSISF 464  
Qy 492 L 492  
Db 465 L 465

RESULT 3  
B82023  
probable periplasmic protein NMA0282 [imported] - Neisseria meningitidis (strain Z2491  
C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: B82023  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; More  
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.  
A:Reference number: AB1775; MUID:20222556; PMID:10761919  
A:Accession: B82023  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-421 <PAR>  
A:Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CA889589.1; PID:g737904  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: NMA0282  
C:Superfamily: hypothetical protein ycdB  
Query Match 7.1%; Score 185; DB 2; Length 421;  
Best Local Similarity 23.2%; Pred. No. 2.2e-06;  
Matches 88; Conservative 41; Mismatches 125; Indels 126; Gaps 17;  
Qy 135 LGITDLDLGDQAPPDGQFA-----DAAN-----LGDGLSQWVAPFTGTTIH 174  
Db 130 LTVTVGVGSSILF-DGRFGLKDKKPIHLQEMROFSDNKLQKSWCDGLSLQICAPTPETCQ 188  
Qy 175 GVFLIGSDQDDFLDQFTDDISSTFGSSITQ--VOALSGSARPFDOAGHEHFGFLDGLISQP 232  
Db 189 AA-----LRDIIKHTVQTAIRWSIDGQKQSEFNAARNLGLFRDGTGNP 235  
Qy 233 SVTGWETTVPQAVVPGIILTRGDGDTGTRPSWALDGSFNAFRHFOOKVPEFNAYTLA 292  
Db 236 KVSQPKTA-----DEVLTGVAANSILDEPEWAKNGSYQAVLRHFRHVEFWDRTP- 285  
Qy 293 NAIPANSAGNLTQEGAEFLGARMFGRWKSGAPIDLAPTDAPALGADPQRNNFDYSDT 352  
Db 286 -----QEQTDIFERRKY-----SGAPMDGKKEADQDPFAKQPEGNTT- 322  
Qy 353 LTDETRCPFGAHVRKTNPRQDLGGP--VDTFHAMRSSIPYGPETSDAELASGVTAQDRGL 410  
Db 323 -----PKDSHIRLANPRD-----PEFLKXHLFRAYS- ----SRGLASSG--QLDVGL 365  
Qy 411 LFVEYQSTIIGNFQFQINWANNANFPFSKPIITPGIEPIIGQTPRTVGGDLPLNQNETF 470  
Db 366 VFVCYQANLADGFIQVQ--NLNG-----EPLEE----- 392  
Qy 471 TVPLFVIPKGGEYFF-LPSI 489  
Db 393 ---YISPFGGGYFFVLPGV 408  
RESULT 4  
F81244  
conserved hypothetical protein NMB0036 [imported] - Neisseria meningitidis (strain MC58  
C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: F81244  
R:Tetelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
rt, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizze, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Frazer, C.M.; Moxon, E.R.; Rappuoli, R.; V.  
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A:Reference number: AB1000; MUID:20175755; PMID:10710307  
A:Accession: F81244  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-421 <TET>  
A:Cross-references: GB:AE002364; GB:AE002098; NID:g725269; PIDN:AAF40507.1; PID:g725252.

A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB0036  
C:Superfamily: hypothetical protein ycdB

Query Match 6.8%; Score 177; DB 2; Length 421;  
Best Local Similarity 23.4%; Pred. No. 8.8e-06;  
Matches 89; Conservative 41; Mismatches 124; Indels 126; Gaps 18;

Y 135 LGITDLDGAPPDGGA-----DAAN-----LGDLSQWVAFGTTH 174  
D 130 LVTYGVSSSLF-DKFLKDKKPHLEQENDFNSNDLQKSKDCGLSDICAFPTTCQ 188  
Y 175 GVFLIGSDDDFLDFTDSSITGSSITQ--VQALSGARPDOAGHEHFGFLDGISQ 232  
D 189 AA-----LRDILIKHTVQAVIRMSIDGMQPKSEPGAMARMLGFRDGTNP 235  
Y 233 SYTGWETTPGQAVVPPGILLTGDDDTGRPSWALDGSFMAFRFQKVPFNATYLA 292  
D 236 KVSDEKTA-----DEVMTGVAANSLDEPMANQSYQAVRLIRHFVEFMDKTP- 285  
Y 293 NAIIPANSAGNLTOGAEFLGARMFGKWSGAPIDLAFTADPALGADPQRNNPDSPT 352  
D 286 -----GEQTIIFGRRY-----SGAPMDGKKEADQDFADPE----- 318  
Y 353 LNDTRCPGCAHVRKTNRPDLGSP--VDTFHMRSSIRPGPETSDAELASGVTADRL 410  
D 319 -GDIT--PDSTIRLANRPD---PEFIKGRLEFRRAVSY---SGLASSG--QLDVGL 365  
Y 411 LFEVYQSIIGNGFRFOQINMANNANFPESKEITPGIEPIIGTTPRTVGLDPLNOMET 470  
D 366 VFVCYQANLADDFIVQ--NLING-----EPLER----- 392  
Y 471 TVPLFVIRPKGEYFF-LPSI 489  
D 393 ----YISPGGGRFVLPGV 408

RESULT 5  
AE0226  
conserved hypothetical protein YP01856 [imported] - Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C>Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Nov-2001  
C:Accession: AE0226  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Dougan, G.;  
11. M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,  
Nature 413, 523-527, 2001  
A>Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AE0226  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-434 <KUR>  
A:Cross-references: GB:AL590842; PIDN:CAC90673.1; PID:g15979878; GSPDB:GN00175  
C:Genetics:  
A:Gene: YP01856  
C:Superfamily: hypothetical protein ycdB

Query Match 5.5%; Score 143.5; DB 2; Length 434;  
Best Local Similarity 22.9%; Pred. No. 0.0033;  
Matches 81; Conservative 38; Mismatches 123; Indels 111; Gaps 16;

Y 145 OPFDGQFADNALGDDLGSOWVAFPTGTTIHGVFLIGSDDDPFDQFDDISSTF---GSS 201  
D 173 RFPNDSLADGLCHGVMLQICANTNEYIHAL-----RDLIKRTPDLISVWKKEGFI 225  
Y 202 ITQVQALSGSARPPQAGHEHFGFLDGISQPSVTGWETTPGQAVVPPGIIITGSDGT 261  
D 226 SAHAASKQDTPINL-----LGFQDGTANPKIS-----NKPILNNVWVSNNAE- 271  
Y 262 GRRPSWALDGSFMAFRFHQKVPFNATYLANAIPANSAGNLTOGAEFLGARMFGPK 321

D 272 ---PAWAVGGSQVVRIRFKV-EFWDRT-----PLGEQ-----TIFGRDK 309  
Y 322 -SGAPIDLAFTADDPALGADPQRNNNPDYSDTLTDRCPGCAHVRKTPRODLCGPDT 380  
D 310 NSGAPLGQHEHDEENYAKDEGK-----VIMQDRIFLANP-----TLET 351  
Y 381 FH-AMRSSIRPGPETSDAELASGVTADRLGFVEYQSIIGNGFRFOQINMANNANFP 438  
D 352 QNNMLRGRYSYSLGVN---SG--QLDMGLLFVCGQSLAQAFITVGBRLNGEALIEY 405  
Y 439 SKRTPGIEPIIGTTPRTVGLDPLNOMETFFVLPVIRPKGEYFF-LPSIS 490  
D 406 VKPI-----GGGYFFLPGVA 421

RESULT 6  
A99787  
hypothetical protein EC61265 [imported] - Escherichia coli (strain O157:H7, substrain R1  
C:Species: Escherichia coli  
C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
C:Accession: A99787  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Iehii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom  
A:Reference number: A99629; MUID:21156231; PMID:11256796  
A:Accession: A99787  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-423 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA034688.1; PID:g13360725; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain R1MD 0509952  
C:Genetics:  
A:Gene: EC61265  
C:Superfamily: hypothetical protein ycdB

Query Match 5.1%; Score 133.5; DB 2; Length 423;  
Best Local Similarity 22.3%; Pred. No. 0.018;  
Matches 81; Conservative 29; Mismatches 119; Indels 135; Gaps 14;

Y 145 OPFDGQFADNALGDDLGSOWVAFPTGTTIHGVFLIGSDQDPL-----DQFDDISSTF 198  
D 162 RFPNDSLADGLCHGVMLQICANTQDVIHALRDIIRTPDLISVWKKEGFISSHAAS 221  
Y 199 GSSITQVQALSGSARPPQAGHEHFGFLDGISQPSVTG-----WETTPGQAVVPP 250  
D 222 KGERPILN-----GFRDGTANPDSONDKLQKXVWVTA----- 256  
Y 251 GILLTRDODDTGTRPSMALDGSFMAFRFHQKVPFNATYLANAIPANSAGNLTOGEGAE 310  
D 257 -----DQGBPAWMTIGSYQAVRLQFRV-EFWDRT-----PKKEQ----- 291  
Y 311 FLGARWFGWK-SGAPIDLAFTADDPALGADPQRNNNPDYSDTLTDRCPGCAHVRKTN 369  
D 292 ---TIFGSDQGTGAPLQNGHEHVDVPRVADPEKG-----IALDSHIRLAN 334  
Y 370 PRODLGAPVDTFHMRSSIRPGPETSDAELASGVTAG--DRLGFVEYQSIIGNGFRFO 426  
D 335 PRT--AESSESLMLRGRYSY-----SLGVTNSGQDMGLFVCTQHDLEKGFLLTV 382  
Y 427 QINMANNANFPESKDIPTGIEPIIGTTPRTVGLDPLNOMETFFVLPVIRPKGEYFF- 485  
D 383 QKRLNGEALIEEVKPI-----GGGYFFLPGVA 405  
Y 486 LPSI 489  
D 406 LRGV 409

RESULT 7  
AB5647  
hypothetical protein ycdB [imported] - Escherichia coli (strain O157:H7, substrain EDL93.  
C:Species: Escherichia coli

[illegible]

A:Gene: DBR0067	
A:Map position: megaplasmid	
A:Genome: plasmid	
A>Note: plasmid MPI	
Query Match	4.9%
Best Local Similarity	23.0%
Score	127.5;
DB 2;	Length 1067;
Pred. No.	0.2;

```

Query Match      4.9%; Score 127.5; DB 2; Length 1067;
Best Local Similarity 23.0%; Pred. No. 0.2;
Matches 121; Conservative 60; Mismatches 171; Indels 175; Gaps 28;

Qy 13 LVGSSSHVNAA---KLGARQRTTFL-----LTNFGQAP 44
      | | | | | : | | | | | : | | |

```

```

Db      55 LSGAGSAIKQASVORLGAQATEVGLSPKSNVYGVYDSSNNTLRMMATFEDVTNTSDQ-P 113
Qy      45 L-PLTCHTSSGANDT-LPLNIOGDLVGMKKQKRFPEFOVNAITSP----- 93
Db      114 LAAPFIPIVDTE-GAGGTTGTTAFKQVR-----YFQSDSASSAPLAFD 157
Qy      94 -----KTALKTVPORITSAILISDPSOCPFAVNLGFSNTG--LCAL-GITDDLGA 144
Db      158 SSNDPRTAVKTOGLDLDSGVVNLPAQLGLA-----GTSHTGMKKGALPAGAT----- 207
Qy      145 QRPDGFADANLGDLSQWVAPFTGTHVFLIGSDODDFLDQFDIDISSTFGSSITQ 204
Db      208 ----GGVTLAASIPAAISSQDNFSS-----DPLVETVADVNPSTLTNLICA 249
Qy      205 VQALS-GSARPPDOAGHEHFGFLDGISQPSVTGWEITVPQOAVVPPGILLTG---RDG 259
Db      250 VQQTGGGDRPALTGAAQ-----TIEGVTVAPER-----LSGFVQEGEG 289
Qy      260 DICTRSMALDSFMAFRHFOQKVEPFENATLANAIPANSQNLTOQEG--APFLAARNF 317
Db      290 IDADRDETTSDGLV-----YQASGC-PALSHGDKVRVSGTVAAYGATQ 334
Qy      318 ----GRWKSGAIDILAPTADDPALGADPQRNNPD-----YSPTLTDETRCPFGAHVRK 367
Db      335 TAPTVTKLLSLGLAL--PPAALMLKPLDKTQGERYEGMRVRFPELTITNNYTRYH---- 388
Qy      368 TNRQDLGCPVVTFFHAMSSIPYGETDAE-----LASGYTAQ--DRGLLVEYQSI 419
Db      389 -----QQLDSNAGRFMFPTNGNASASOSTITLLDDGSAQNPDLNLVLSAERTRT 440
Qy      420 GNGFRFQOIMNANNAFPFSKPTPGIEPIIGOTT-----PRTVGG 460
Db      441 GQTVVGLSGVHMSVANQPLRP--EGAVEFVSANSRASNMLPHVYGG 485

```

```

RESULT 10
AH0367
conserved hypothetical protein YPO3025 [imported] - Yersinia pestis (strain CO92)
C/Species: Yersinia pestis
C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #ext_change 03-Nov-2001
C/Accession: AH0367
R/Archival: J.; Wren, B.W.; Thomson, N.R.; Tibbells, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-farrage, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
I., M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,
Nature 413, 523-527, 2001
A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A/Reference number: AB0001; MUID:21470413; PMID:11586360
A/Accession: AH0367
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-299 <KOR>
A/Cross-references: GB:AL590842; PIDN:CA092267.1; PID:G15980978; GSPDB:GN00175
A/Genes: YPO3025
C/Superfamily: Escherichia coli hypothetical protein b2431

```

```

Query Match          4.9%  Score 126.5; DB 2; Length 299;
Best Local Similarity 21.7%; Pred. No. 0.038;
Matches 91; Conservative 45; Mismatches 109; Indels 175; Gaps 23;

```

```

Qy      109 AILISDPSOCPFAVNLGFSN--TGLQALGITDDLGAKQFPD-----GQFADANLGDLS 162
Db      16 AIFMEAKVQGEIDAIRLGCKKFCQSLQEL-----QQQFDEHLGAVIVFGSNVWHIDS 68
Qy      163 -----QWVAPFTGTHVFLIGSDODEL---DQFTDISSTGSSITVOVALSGSARP 214
Db      69 NGGAAELKRFVPL--GKGLAPATQRLDLHITQSLKODINFTLAQAAV---AAGSATA 122
Qy      215 FDOAGH-----EHFGFLDGISQPSVTGWEITVPQOAVVPPGILLTGSDPTGRRS 266
Db      123 VEBETHGFRVVEERDFGTIDGTEPN-----QGDKRPEVAVIADG-----EEDAG--- 167
Qy      267 WALDGSFMAFRHFO-----QKVPFENATLANAIPANSAGNLQOGAEFLGARMGR 320

```

```

Db      168 ----GSYLVQKRYEHNNLMKQRIPE-----NEGE-----KIORT 198
Qy      321 KSGAPIDLAPTRADDPALGADPQRNNNFDYSOTLTDETRCPFGAHVRKTPRODLOGPVD 380
Db      199 K-----LDSQELPSD--QR-----PDSHSRVDLKENGKG--- 227
Qy      381 FHMRSSIPYGPETSDALAGVTAODRGLLFVEY-----QSTINGGFPFOQIMNAN 432
Db      228 EKLIRQSLPYG-----LASG-----KHGLYFLAYCARLHINEQQLS--MFGIDGKH 273
Qy      433 NANFPFSKPTPGIEPIIGOTTPTTVGGDLPLNONTFTVPLFVIEKGEYFFLDSIAL 492
Db      274 DQLRRSKPEVT-----GSYVFAPSLTAL 296

```

```

RESULT 11
B86748
hypothetical protein ykbc [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C/Species: Lactococcus lactis subsp. lactis
C/Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #ext_change 03-Aug-2001
C/Accession: B86748
R/Bolotin, A.; Winkler, P.; Mager, S.; Jallion, O.; Malarme, K.; Weisenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A/Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp
A/Reference number: AB6625; MUID:21235186; PMID:11337471
A/Accession: B86748
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1093 <STO>
A/Cross-references: GB:AA005176; PID:G12723930; PIDN:AAK05084.1; GSPDB:GN00146
A/Experimental source: strain IL1403
A/Genes: ykbc

```

```

Query Match          4.7%  Score 121.5; DB 2; Length 1093;
Best Local Similarity 21.3%; Pred. No. 0.59;
Matches 123; Conservative 64; Mismatches 228; Indels 163; Gaps 27;

```

```

Qy      17 SSVNNAKIKGAGQRTTP-LTNFPQAPLPTLQ--HTTSGANDTILPLNNIOGILVG 74
Db      62 SPVNSSLAASSTEEBAVVSNNMENTROVSSNITLITSSVVDNPSFGNSIALS 121
Qy      75 MKKQKREVFVQVNDATSPKALKTVYVPRITSAILISDPSOCPFAVNLGFSN---- 129
Db      122 LTKASRESSINPASTIANETTTADVTQTATQATQSDPTTS-LSSSOGKPTNPSQS 180
Qy      130 ----TGLQALGITD--LGAQFPDQGFADANLGDLSQWVAPFTGTHVFLIGSD 182
Db      181 KSTETINIQVTAGVDMNATGSAVF-----DGVNI-----TLQS----- 213
Qy      183 QDDFLDQFTDDISSTFGSSITVOVALSGSARPPDQAGHEHFGFLDGISQPSVTGWEITV 242
Db      214 -KDITDNLDPGSLHWESEQTVIAIKGTAT--COLNQENFGVDDGILIPATT---YTNW 267
Qy      243 PQGA--VVPFGILLTGROGD-----TGTRPSW-ALDGSFMAFRHFOQKVEFNATLAN 293
Db      268 DQSGSITTVGSLSGDLIDMTIYVASSDKDSQWAMEGA-----EGIPQGLTTGEO 319
Qy      294 AIPANSAGN--LTQDEGAFLGARMFGKWSGADIDLAPTAD-----DPAIGAD-- 340
Db      320 NI--AESGNSIVCLVNGANAL-SLIYQIVKADTTTEVPVVASFTTDDIDNAQGVQTNLAN 377
Qy      341 ----PQRNNNFDYSDTLTDET-----RCQPFARH----- 366
Db      378 LVTLIPOTTNLKODGDTIYDASPNYPGLDGVASLPYGGYIAGAFVSEFYNNYAPAPERA 437
Qy      367 -----KTNRQDLGCPVDTFFHAMSSI-----PYGPTSQALASGYTAQDRGL 410
Db      438 DSYFPAQGVRYDLFGSALQAH-MNTQIRQNFYVYVDEFGKHIOETDHLGFTGQDVM 496
Qy      411 LVEYQSIINGNF-----RFOQIMNANANP-----FSKPTTP 444

```

Db 497 ---PIPTIKGFVNLNTENDASKNNPVINLYNHNLTYYGNHNNNIYYOGTAYTPSFYI 553

QY 445 GIEPIIGTTPRTVGGDLPLNQNTEFTVPLFVLPKGE 482

Db 554 GYQNI---GNPEASITYTPVANGKASSVTPLIMAGOK 588

RESULT 12

T50585

probable membrane protein [imported] - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C>Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 28-Jul-2000

C:Accession: T50585

R:Reisenbach, M.; Kleser, H.M.; Denapaita, D.; Eichner, A.; Cullum, J.; Kinashi, H.; Hopm

Mol. Microbiol. 21, 77-96, 1996

A:Title: A set of ordered cosmid and a detailed genetic and physical map for the 8 Mb S

A:Reference number: 220556; MUID:97000361; PMID:8843436

A:Accession: T50585

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-420 <RED>

A:Cross-references: EMBL:AL133220; PIDN: CAB61722.1

A:Experimental source: strain A3(2)

C:Genetics:

A:Note: SCC75A.22

C:Superfamily: hypothetical protein ycdB

Query Match 4.6%; Score 119.5; DB 2; Length 420;

Best Local Similarity 21.98; Pred. No. 0.21;

Matches 81; Conservative 35; Mismatches 121; Indels 133; Gaps 17;

QY 89 DATSFKTALKTYVPORITSAAILISDPQPLAFVNLGFSNTGLQALGTTDLGDA---- 144

Db 125 DOTGEALGLK---PSRLT-----LTIGFGLSLFTRFGLADLRPEALADL 165

QY 145 -QFPDQGFADANLGDLS-QWVAPFTGTTIHGV---FLIGSDDDFLDOFTDDISSTFG 199

Db 166 PKFP-GDNLDRARSGDLCVQACADDQVAVHAINRLRIG-----FG 207

QY 200 SSITOVQAL----SGSARPPQAGHEFGFDGISQPSVTGWTTFVFGQAVVPPGIILT 255

Db 208 KVVVRWSQLGFKTSSTTPDSQTPERNLLGPKDTRN-----IA 245

QY 256 GRDGTGTRPSWALD-----GSFMRHFOOKVPEFNAYTLANAIPANSAGNLTCQ 306

Db 246 GTEKRLDRFVVAEKDGTPTMTGSGYLVARRIMHLETWD-----RASLQEQ 293

QY 307 EGAEFLGARMEGRWK-SCAPIDLAPTADDPALGA-DPQRNNNFYSOTLTDTETRCPPGAH 364

Db 294 ED-----VFGRDKGEGAPVGKAKERDEPFLKMKPD-----AH 326

QY 365 VRKTNPRDLGGPVDTHAMRSSIPYGPETSDAELASGVTAQDRGLLFVEYQSIINGFR 424

Db 327 VRLHPDSNGGNTL-----LRGYSFTDGT-----DGLGRDLNAGLFFLAYORDTGTGV 375

QY 425 FOQINWANNA 434

Db 376 PVQRNLATDA 385

RESULT 13

AB3412

hypothetical protein PA1874 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: AB3412

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

o; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: AB2950; MUID:20437337; PMID:10984043

A:Accession: AB3412

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2468 <STO>

A:Cross-references: GB:AE004613; GB:AE004091; NID:g9947856; PIDN:AAG05263.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA1874

Query Match 4.6%; Score 118.5; DB 2; Length 2468;

Best Local Similarity 21.1%; Pred. No. 3.2;

Matches 115; Conservative 57; Mismatches 218; Indels 155; Gaps 28;

QY 21 NAAKL-GARQRTTPLLNTNPGQAPLPTLTQHTTESGANDTILPLNNTQGLIYGMKKQK 79

Db 1088 NGSSLGTAEPGSGVIITLNGN-----PLAETVADGSGNWTTPSTPIANGTVVNVAAQ- 1142

QY 80 ERFVFPQVNDATSFKTALKTYVPORITSAAILISDPQOPLAFVN--LQFSNTGLQALGI 137

Db 1143 -----DASG-----NSPPATVTVDSAPPVNPVNSGVVSGTAEAGA 1182

QY 138 TDDLGA-QFPDQGFADANLGDLSQWVAPET-GTTI-HGVFLIGSDDDFLDOFTDDI 194

Db 1183 TVTLTDAGGNPIQGV-----ADSGNW--SFTPGPLANGTVIVAT-----ATDPT 1227

QY 195 SSTFGSSITVOALSGSARPPDQAGHEFGFDGISQPSVTGWTTFVFGQAVVPPGIIL 254

Db 1228 GNTGPOAATTVDVAPPAPVID-----PS-----NGTTISGTAEAGAKVIL 1268

QY 255 TCRDGTGTRPSWALDGSFAPRHFQOKVPERNAYTLANAIPANSAGNLTOEGAEFL-- 312

Db 1269 T--DNGNPIGETTADGS--GNWSFTPGTPLANG--TVNNAVAQDPAGN--TGPQSGTTVDA 1322

QY 313 -----GARMEGRWKSGAPIDLAPTADDPALGADPQRNNNFYSOTLTDTETRCP 360

Db 1323 VAPNTFVVNPNSGNLLNGTAEPGSTVTLTDGNGNPIGQTTADGSGNWSF----TPGSQLP 1378

QY 361 GAHVRYKTNPRDLGG-----PVDTFHAMRSSIPYGPET--SDAELASGVTAQDRGLLFVE 414

Db 1379 NGTVNVVT--ASDAAGNTSLPAT--TVDSULPSIPQVDPDSNGSVISGTADAGNTIITD 1434

QY 415 YOSTIGNGFRPQOI-----NWANNANFPFSS-----KP 441

Db 1435 -----GNGNPIGQVTADGSGNWSFTPGIPLDQGTVVNVARSNVSAPAVITVDGVAP 1489

QY 442 ITPGTEPI-----IGOTTPRTVGGDLPLNQNTEFTVPLFVLPKGEYFFLPSISA 491

Db 1490 AAPVLDPSNGTEISGTAEAGATVILTDGNGNPIGQ-----ATADGSGNWTTFPSTPL 1541

QY 492 LTATI 496

Db 1542 ANGTV 1546

RESULT 14

D64962

probable membrane protein bl978 - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C>Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002

C:Accession: D64962

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: D64962

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2383 <BLAT>

A:Cross-references: GB:AE000289; GB:U00096; NID:gl788285; PIDN:AAC75042.1; PID:gl788288,

A:Experimental source: strain K-12, substrain MG1655

C:Keywords: nucleotide binding; P-loop; transmembrane protein

F:54-70/Domain: transmembrane #status predicted <TM>

F:1564-1571/Region: nucleotide-binding motif A (P-loop)

Query Match 4.5%: Score 118; DB 2; Length 2383;  
 Best Local Similarity 19.5%; Pred. No.3.3; 160; Indels 178; Gaps 19;  
 Matches 95; Conservative 54; Mismatches 160

```

QY 106 TSAAILIDSPSOQPLAFVNLGFS-----NTGL---QALG-----136
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1086 TLTATVXDPNSNHPAGITVFTMQDVAAANFTLENNGIAITQANGSAHYTLKGKAGTH 1145
QY 137 -TTDDLGDAAFPDGG-----PADAAANLGDGLSQWAFPTGTHGVFLIGSDDDFLDFT 191
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1146 TTTATLGNNTSDSQPTFVADKSAQVVLQISKDEITGNGVSATLTATVKDQF-DNEV 1204
QY 192 DDISTFGSSITQVALSGSARPPDQAGHEHGFLDGISQPSVTGMETTVFPQAVVPPG 251
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1205 NNLPVTFSSASGLTLTPGVSNTNE-----SGIAQATLAG--VAFGEKTVTAS 1250
QY 252 ILLTGDDGTGRPSMALDGSFMAFRHQKVPFPMAYTLANAIIPANSAGNLTOEGAF 311
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1251 LANNAS-----DNKTHFIDGTAAKI-----1273
QY 312 LGARFGRWKSAPIDLAFTADDPALGADPQRNNPDFSDTLTDETRCPF-----361
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1274 -----IELAPVPDSIIAGT-PQNSSGSVITATVVDNNGFPVKGVTVFTSN 1318
QY 362 GAHVKTNPQ---DIGSPVDTFHAMRSSIPYG--PETSDAELASG-----402
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1319 AATAEWNGGQAVTNEQKATVYTNTRSSIESGARDPTVEASLENGSSTLSTINVAD 1378
QY 403 -----VTAQDRGLLFVEYQSIIGNGFRFOQIMNANNANFPSPKIPGIE 447
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1379 ASFAHLTLQALFDTVSAGETTSYIEVKDNYGNGVPOQEVTLVSVP---SEGVT 1434
QY 448 PIIGOTPRTVGGLDPLMNQETFTVPLFVIRKGEY-----FELPSISA 491
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1435 AI--YTT-----NHDGNFYAS-FTATKAGVGLTATLENGSMQGTVTYVENVAN 1481
QY 492 LTATIAA 498
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1482 AETTLAA 1488

```

## RESULT 15

E85822  
 probable invasin 23135 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)  
 C:Species: Escherichia coli  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C:Accession: E85822  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Miller, L.; Grodzbeck, E.J.; Davis, N.W.; Dim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: A85480; MUID:21074935; PMID:11206551  
 A:Accession: E85822  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2660 A:STO>  
 A:Cross-References: GB:AE005174; MTD:G12516151; PIDN:AAG57041.1; GSPDB:GN00145; UWGP:Z31  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetic8:  
 A:Gene: 23135

Query Match 4.5%: Score 116; DB 2; Length 2660;  
 Best Local Similarity 18.2%; Pred. No.5.5;  
 Matches 114; Conservative 77; Mismatches 21; Indels 224; Gaps 26;

```

QY 11 AVLVGSSSHVNAKLGARQRTTPLL-----TNFGQADPLTLTQHTTE 54
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1245 AAVVLOTSTKAEITIGVGVDETTLTATVXDPNNAVKDLQVTFSTN-PADTQLSQSKSTND 1303
QY 55 SGAND-----TTLPLNNIGDILVGMKKKQKRFVFPQVNDATSPKLTXTYVPORTS-- 107
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1304 SGVAEYTFPKGTVGVHTATLTPNGNNDTK---IVNTAPASNAQVTLNIPAQVVTNNS 1360

```

```

QY 108 ----AAIILSPSOQPLAFVNLGFS-----NTGL---QALG-----136
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1361 DSVQLTATVXDPNSNHPAGITVFTMQDVAAANFTLENNGIAITQANGSAHYTLKGKAG 1420
QY 137 -TTDDLGDAAFPDGGPADAAANLGDGLSQWAFPTGTHGVFLIGSDDDFL 187
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1421 TTTATLGNNTSDSQ--PVTFAADKTSALVVLQISKDEITGNGVSATLTATVKDQF- 1477
QY 188 DQFTDDISTFGSSITQVALSGSARPPDQAGHEHGFLDGISQPSVTGMETTVFPQAV 247
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1478 DNEVNNLPVTFSSASGLTLTPGVSNTNE-----SGIAQATLAG-----1516
QY 248 VPPGIIITGRDGTGRPSMALDGSFMAFRHQKVPFPMAYTLANAIIPANSAGNLTOE 307
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1517 -----VAFGEQTVTASLANNGA---SDNK 1537
QY 308 GAFLGARFGRWKSAPIDLAFTADDPALGADPQRNNPDFSDTLTDETRCPF-----361
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1538 TVHFIGDT---AAAKIILTPVPDSIIAGT-PQNSSGSVITATVVDNNGFPVKGVTVN 1591
QY 362 ---GAHVKTNPQ---DIGSPVDTFHAMRSSIPYG--PETSDAELASG-----402
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1592 FTSNATAEMTGGQAVTNEQKATVYTNTRSSIESGARDPTVEASLENGSSTLSTIN 1651
QY 403 -----VTAQDRGLLFVEYQSIIGNGFRFOQIMNANNANFPSPKIP 443
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1652 VANADSTAHLTLQALFDTVSAGDTNLYIEVKDNYGNGVPOQEVTLVSVP---SEGVT 1707
QY 444 PGIEPI-----IGOTPRTVGGLDPLMNQETFTVPLF-----475
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1708 PSNNAIYTNTHDGNFYASFTAXKAGVQVT-ATLENNDSMQGTVTY-VENVANAETSLAA 1765
QY 476 ----VIRKGEYFELPSISALTATIA 497
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1766 SKDPVIANND-----LTTLTATVA 1785

```

Search completed: November 27, 2002, 13:48:18  
 Job time : 29 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 27, 2002, 13:45:31 : Search time 14 Seconds  
(Without alignment)  
1475.372 Million cell updates/sec

Title: US-09-926-084-7

Sequence: 1 MDLSLFVSAVALVGSSESHV.....KGGYFFLPISALTATIAA 498

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	130.5	5.0	423	1 YCDB_ECOLI	P31545 escherichia
2	118	4.5	2358	1 YEBJ_ECOLI	P76347 escherichia
3	114	4.4	2660	1 YEBJ_ECO57	O88XV7 escherichia
4	108.5	4.2	953	1 LMA3_PASHA	P55116 paestherella
5	108	4.2	647	1 NMMH_MICVI	Q02834 micromonos
6	107	4.1	3562	1 PCGV_CHICK	Q90953 gallus gall
7	105	4.0	1229	1 N121_HUMAN	O9Y2N3 homo sapien
8	104.5	4.0	430	1 N042_YEAST	P49686 saccharomyc
9	104	4.0	603	1 BPRV_BACNO	P42779 bacteroides
10	104	4.0	753	1 CATE_ECOLI	P21179 escherichia
11	103.5	4.0	710	1 FOXA_YEREN	Q01674 yereshia en
12	103	4.0	1076	1 N0P1_YEAST	P03676 saccharomyc
13	102.5	3.9	1015	1 POL_HVIBR	P13126 deinnococc
14	102.5	3.9	1036	1 HP12_DEIRA	P32521 deinnococc
15	101.5	3.9	1480	1 PANT1_YEAST	O67988 rhodococcu
16	101	3.9	252	1 CLCD_RHOOP	O09838 schizosacc
17	101	3.9	681	1 YADD_SCHPO	O9Y272 deinnococc
18	101	3.9	890	1 SYA_DEIRA	P32790 saccharomyc
19	101	3.9	1244	1 SYA1_YEAST	O43511 homo sapien
20	100.5	3.9	1856	1 MGA_HUMAN	Q9HQZ2 halobacteri
21	100	3.8	600	1 IP2P_HAANI	P04589 human immu
22	100	3.8	1002	1 POL_HVIEI	P76115 escherichia
23	99.5	3.8	700	1 YNCD_ECOLI	P36712 human adeno
24	99	3.8	582	1 HEX3_ADEI2	P03169 human immu
25	98.5	3.8	1003	1 POL_HVIA2	P11079 reovirus (t
26	97.5	3.7	1003	1 POL_HVIB2	P04585 human immu
27	97.5	3.7	2124	1 POL_HVIB2	P07897 rattus norv
28	97.5	3.7	2124	1 POL_HVIB2	P18802 human immu
29	96	3.7	1002	1 POL_HVIB2	P20892 human immu
30	95.5	3.7	1003	1 POL_HVIB2	P52948 homo sapien
31	95	3.7	937	1 N098_HUMAN	P12497 human immu
32	95	3.7	1003	1 POL_HVIB2	P33666 escherichia
33	95	3.7	2003	1 YDBA_ECOLI	

34	94.5	3.6	735	1	DBB4_MOUSE	P51660 mus musculu
35	94.5	3.6	937	1	N098_PAT	P49793 rattus norv
36	94.5	3.6	2090	1	N214_HUMAN	P33658 homo sapien
37	94	3.6	1002	1	POL_HVIBR	P05959 human immu
38	94	3.6	1089	1	MLB_MYCTU	O07800 mycobacteri
39	93	3.6	1002	1	POL_HVIB2	P12499 human immu
40	93	3.6	1189	1	ITAH_HUMAN	Q9UKX5 homo sapien
41	92.5	3.6	959	1	N100_YEAST	O02629 saccharomyc
42	92.5	3.6	978	1	PM14_CHLUP	O92895 chlamydia p
43	92.5	3.6	995	1	Y109_YEAST	P40442 saccharomyc
44	92.5	3.6	1006	1	POL_HVIBN	P05961 human immu
45	92.5	3.6	1015	1	POL_HVIB5	P04587 human immu

## ALIGNMENTS

RESULT 1  
YCDB\_ECOLI STANDARD; PRT; 423 AA.  
AC P31545; P75903;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein ycdb precursor.  
GN YCDB\_OR\_B1019  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blatner F.R., Plumett G., Ili, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shaoy Y.;  
RT "The complete genome sequence of Escherichia coli K-12";  
RL Science 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97061202; PubMed=8905232;  
RA Oshima T., Alba H., Baba T., Fujita K., Hayaishi K., Honjo A.,  
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
RA Yano M., Horinouchi T.;  
RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 12.7-28.0 min region on the linkage map";  
RL DNA Res. 3:137-155(1996).  
RN [3]  
RP SEQUENCE OF 196-423 FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=93186698; PubMed=844794;  
RA Kim S.-K., Makino K., Amemura M., Shingawa H., Nakata A.,  
RT "Molecular analysis of the pho gene, belonging to the phosphate  
regulon in Escherichia coli";  
RL J. Bacteriol. 175:1316-1324(1993).  
RN [4]  
RP SIMILARITY: TO B.SUBTILIS YWBN (IPA-29D).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DR EMBL; AE000203; AAC74104.1; -  
DR EMBL; D90738; BAA35796.1; -

```
DR EMBL; D90739; BAA35800.1; -
DR EMBL; D10391; BAA01229.1; -
DR PIR; A47065; A47065.
DR EcoGene; EG11735; ycdB.
DR PROSITE; PS00013; PROKAR LIPOPROTEIN; UNKNOWN 1.
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 325
FT CHAIN 36 423
FT SEQUENCE 423 AA; 46754 MW; 65D381F829DB2570 CRC64;

Query Match 5.0%; Score 130.5; DB 1; Length 423;
Best Local Similarity 22.3%; Pred. No. 0.021;
Matches 81; Conservative 29; Mismatches 119; Indels 135; Gaps 14;

QY 145 QPDCQFADANLGGDLSQWVAPPTGTHGVFLGSDQDDPL-----DQTDIDISSTF 198
DB 162 RPNDSLDAALCHGVLLQICANTQDVIHALRDIHKPTDLLSVRWKREGFISDHAARS 221
QY 199 GSSITQVQALSGSARPPDQAGHEHFGFLDGISQPSVTG-----WETTVPPGQAVVPP 250
DB 222 KKETPINLL-----GFKDGTANPDSONDKLMQKVVWTA----- 256
QY 251 GIILTRGDGTGTRPSWALDGSFMAFRHFOOKVPEFNAYTLANAI PANSAGNLTQOEGAE 310
DB 257 -----DQOEPAMTIGGSQAVRLIQFRV-EFWDR-----PLKEQQ----- 291
QY 311 FLGARMGRWK-SGAPIDLAPTADDPALGADPQNNNNFYSDTLTDTETRCPEGAHVTKTN 369
DB 292 ----TIFGRDQKTCAPLQMOHEHDVDPYASDPGK-----VIALDSHIRLAN 334
QY 370 PRQDLGGPVDTFHMRSSIPYGPRTSDAEIASGVTAQ----DRGLLEVEVQSIIGNPRFQ 426
DB 335 PRT-----AESSESLMRGYS-----SLGVTNSGQDLMGLLFVCYQHDLEKGLTV 382
QY 427 QINWANNANFPSPKIPITPGIEPIIGTTPRTVGLDPLNQNETFTVPLFVIPKGEYFP- 485
DB 383 OKRLANGEALEEYVKPI-----GGGYFFA 405
QY 486 LPSI 489
DB 406 LPGV 409

RESULT 2
YEEJ ECOLI STANDARD; PRT; 2358 AA.
AC P76347; P94750;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yeeJ.
GN YEEJ OR B1978.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R. Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasei H., Kimura S., Kitagawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
```

```

QY 403 -----VTAODRGLLFEVEYOSIIGNGFRFOQIMANNANPPFSKPITPTGIE 447
DB 1354 ASTAHLTLLOALFTVTSAGETTSLEYKXNGVNGVPOQEVTLTSSP-----SEGVTSPNN 1409
QY 448 PIIIGQTPRTVGLDPLNQNETFVPLVPIKGEY-----FELPSISA 491
DB 1410 AI--YTT-----NHDSNFPYAS--FTATKAGVLTUATLENGDSMOQVTVYENNVAN 1456
QY 492 LTAITAA 498
DB 1457 AETITLAA 1463

RESULT 3
YEST ECOS7
ID YEEST ECOS7 STANDARD: PRT; 2660 AA.
AC 08X8V7: 08X2C0: 08X2B9:
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Hypothetical protein yeast.
GN 23135 OR ECS2775/ECS2776.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
NCBI_Taxid=83334;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; Pubmed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rosal D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller B.,
RA Godbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potemousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RT Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N. A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; Pubmed=11286796;
RA Hayashi T., Makino K., Chnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
RA "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RN [3]
CC - DNA Res. 8:11-22(2001).
CC - SIMILARITY: CONTRAINS 16 BIG-1 DOMAINS.
CC - SIMILARITY: BELONGS TO THE INTIMIN/INVASIN FAMILY.
CC - CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 1315.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE005423; AAG57041.1; -
CC EMBL: AP002559; BAB36198.1; ALT_FRAME.
CC EMBL: AP002559; BAB36199.1; ALT_FRAME.
CC InterPro: IPR003344; BIG-1.
CC InterPro: IPR000601; PKD_domain.
CC Hypothetical protein; Repeat; Complete proteome.
CC DOMAIN 738 834 BIG-1.1.
CC DOMAIN 840 929 BIG-1.2.
CC DOMAIN 931 1033 BIG-1.3.
CC DOMAIN 1042 1132 BIG-1.4.

```

```

FT DOMAIN 1134 1236 BIG-1.5.
FT DOMAIN 1245 1335 BIG-1.6.
FT DOMAIN 1337 1439 BIG-1.7.
FT DOMAIN 1448 1539 BIG-1.8.
FT DOMAIN 1548 1652 BIG-1.9.
FT DOMAIN 1653 1750 BIG-1.10.
FT DOMAIN 1751 1855 BIG-1.11.
FT DOMAIN 1856 1957 BIG-1.12.
FT DOMAIN 1963 2056 BIG-1.13.
FT DOMAIN 2065 2156 BIG-1.14.
FT DOMAIN 2157 2252 BIG-1.15.
FT DOMAIN 2254 2355 BIG-1.16.
SQ SEQUENCE 2660 AA; 280062 MW; 01EB92A08F5C09D2 CRC64;

Query Match 4.4; Score 114; DB 1; Length 2660;
Best Local Similarity 18.5; Pred. No. 4.1;
Matches 110; Conservative 72; Mismatches 207; Indels 204; Gaps 25;

QY 11 AYLVGSSSHVNAKAGANQRTTPLL-----TNFGQAPLPLTQHTTE 54
DB 1245 AAVVLQTSKAEIIGNGVDETLTATVPDPAWVKDLYTFSTN--PADTQLSQSKSNTND 1303
QY 55 SGAND-----TILPANIQDILVGMKKOKERFVFOVNDATSFKTLKTVYVORTS-- 107
DB 1304 SGVAEYTFKGYLVGHTAEATLPRGNNDTK---IVNIAPDASNAOVTLNIIPAQOVTYNN 1360
QY 108 ---AAILSDPSQOPLAFVNLGFS-----NTGL---QALG----- 136
DB 1361 DSVQLTATVKOPSPNHPVAGITVNTFMPDVAANFTLENNGIAITQANGEAHYTLKKKXAG 1420
QY 137 ---ITDGLDGAQFPFGGCPADAANIGDDLSQVVA-----PFGTTHGVFLIGSDDDPL 187
DB 1421 THVYVNTLSNNNTSSO---PVTFAADKTSALVYLQISKNEITNGVDSATLTAIVKDOF- 1477
QY 188 DOFTDISFSSITVOALSGSARPPDQAGHEHFGFLDGSQSPSVTGMETTVPPGQAV 247
DB 1478 DNEVNNLPVTFSTASSGLTLTPGGSNTNE-----SGIAQATLAC----- 1516
QY 248 VPPGILLTRGDGDTGTRSMALDGSFMAFRHQKVEFNAVYTLAAIPANSAGNLTQOE 307
DB 1517 -----VAFSEQVTVTASLANNA-----SDMK 1537
QY 308 GAEFLGARMFGKSGADPDLAFTADPALGADPORNPNFVSDTLTDETRCF----- 361
DB 1538 TVHFIQDT-----AAAKIIEFLPVPDIIAGT--PQNSGGSVITATVVDNNGFPVKGTVN 1591
QY 362 ---GAHVAKTNPRC---DLGGPVDPFHAMRSSIPYG--PETSDELASG----- 402
DB 1592 FTSNATAEMTNGQAVNBOGKATVYTNTRSSIBSGARPDVBSLNGSSTLSTSN 1651
QY 403 -----VTAODRGLLFEVEYOSIIGNGFRFOQIMANNANPPFSKPIT 443
DB 1652 VNADASTAHLTLLOALFTVTSAGETTSLEYKXNGVNGVPOQEVTLTSSP-----SEGVT 1707
QY 444 PGIEPIIGQTPRTVGLDPLNQNETFVPLVPIKGEYFELPSISALTPT 496
DB 1708 ESNAL--YTT-----NHDSNFPYAS--FTATKAGV-----QVATIL 1740

RESULT 4
ID LKAS3 PASHA STANDARD: PRT; 953 AA.
AC P55116;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Leukotoxin from serotype 73.
GN LKTA.
OS Pasteurella haemolytica.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Mannheimia.
OX NCBI_Taxid=75985;
RN [1]

```



DR InterPro: IPR000421; FAS8\_C.  
 DR InterPro: IPR002860; GH\_BNR.  
 DR Pfam: PF02012; BNR 5.  
 DR SMART: SM00231; FAS8C; 1.  
 KM Hydrolase; Glycosidase; Signal; Repeat: 3D-structure.  
 FT SIGNAL 1 37  
 FT CHAIN 38 647  
 FT ACT\_SITE 260 260 NUCLEOPHILE.  
 FT ACT\_SITE 370 370 NUCLEOPHILE.  
 FT REPEAT 102 113 BNR 1.  
 FT REPEAT 175 186 BNR 2.  
 FT REPEAT 239 250 BNR 3.  
 FT REPEAT 287 298 BNR 4.  
 FT REPEAT 348 359 BNR 5.  
 SQ SEQUENCE 647 AA; 68830 MW; DCCIFESHC935B8AD CRC64;  
 Query Match 4.2%; Score 108; DB 1; Length 647;  
 Best Local Similarity 20.1%; Pred No.1.7; Indels 220; Gaps 23;  
 Matches 110; Conservative 48; Mismatches 169; Indels 220; Gaps 23;  
 QY 79 KERFFVFOVNDATSKFKALKTVFQRTISAAILIDSPSQPL-----AFVN--LGFSNTG 131  
 DB 12 RRRVAVFLAPALAAATVAGASPAQALAGAV-----PGGEPLEYEDLDLVNGREGFPNRY 68  
 QY 132 LQALGITDGLGAGFPDQGFADANLGDGLSQWVAPFTGTHGVFLISDQDDFLDQFT 191  
 DB 69 IRLALVTPD-----GDLASVGRFRTGIDAFG-----FNSILQKRS 104  
 QY 192 DDISTFGSSITVOVALSGSARPFDDQAGHEHGFPLDISPS-VTGMET-TVF----- 242  
 DB 105 TGGCGTWGEO-QVVSAGQTAP-----IKGFSPPSYLVNRETGTFINFFVVSQ 151  
 QY 243 -----PGQAVNPPGI-----ILTGRDG-----DGTSPSMALDGSFMA 275  
 DB 152 RGFAGSRGRTDPAPPNVLHANVATSDGILTWGHRITTDITPDDQMSRFPASGEIG 211  
 QY 276 FHHFOQKVEFNAAVYLLANAIPANSAGNITQEGAEFLGAMFGR-WKSGAFIDLAFTADD 334  
 DB 212 LRYGPHAGRLIOQVYTIINAGAFQVSVYSDD-----HGRTRAG----- 251  
 QY 335 PALGADPOKNNNFYSD-----TLTDERCPFGAHVKRKNPRODGLSPVDTFHAMRSIP 389  
 DB 252 EAVGVGMENKTVESLDGRVLNSRDSKRSQY---RKAVSTD-GS-----HS----- 295  
 QY 390 YGEETSDAEI-----ASCVTAAQDRGLLF----- 412  
 DB 296 YGVFTIDRLDPPTNNAIIRAFPDAPAGSARAKVLLFNSNASTSRSGQTIKMSCDGQ 355  
 QY 413 -----VEYQSII-----GNGFRPQGIN--WANNANFPSPSKP---I 442  
 DB 356 TWPVSKVFPQSGMSITLALPDGTYGILYEPGTHYANFNALMAGIGCAPPTIDVAL 415  
 QY 443 TPGIEBII-----GGTTPR-----TVGGLDPLNQN-----TFVPLFV 476  
 DB 416 EFGQVTVVAVATNGSIAVPKPSIQLDASPDQVQGSVEPLMPGRQAKGVITIVPAGT 475  
 QY 477 IPKGEY 483  
 DB 476 TP--GRY 480  
 RESULT 6  
 PGCV CHICK STANDARD; PRT; 3562 AA.  
 AC 090953; 090945; 1  
 DT 01-NOV-1997 (Rel. 35, Last Sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last Sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Versican core protein precursor (Large fibroblast proteoglycan)  
 GN (Chondroitin sulfate proteoglycan core protein 2) (PG-M).  
 GN CPBG2.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus  
 OC NCBI\_Taxid=9031;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=white leghorn; TISSUE=Limb bud;  
 RX MEDLINE=93300846; PubMed=8314802;  
 RA Shimomura T., Nishida Y., Ito K., Kimata K.;  
 RT "cDNA cloning of PG-M, a large chondroitin sulfate proteoglycan  
 RT expressed during chondrogenesis in chick limb buds. Alternative  
 RT spliced multiforms of PG-M and their relationships to versican";  
 RL J. Biol. Chem. 268:14461-14469(1993).  
 CC - FUNCTION: May play a role in intercellular signaling and in  
 CC connecting cells with the extracellular matrix. May take part in  
 CC the regulation of cell motility, growth and differentiation. Binds  
 CC hyaluronic acid.  
 CC - SUBCELLULAR LOCATION: Secreted; extracellular matrix.  
 CC - ALTERNATIVE PRODUCTS: At least 2 isoforms; V0 (shown here) and  
 CC V1; are produced by alternative splicing.  
 CC - TISSUE SPECIFICITY: Prechondrogenic condensation area of  
 CC developing limb buds.  
 CC - DEVELOPMENTAL STAGE: Disappears after the cartilage development  
 CC (by similarity).  
 CC - SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
 CC - SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.  
 CC - SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC - SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X60226; CAA42787.1; -;  
 DR EMBL; D13542; BAA02742.1; -;  
 DR HSSP; P00740; 1EDM.  
 DR InterPro: IPR000152; Asx\_hydroxyl.  
 DR InterPro: IPR000561; EGF-like.  
 DR InterPro: IPR000742; EGF\_2.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR003599; 1g.  
 DR InterPro: IPR003006; 1g\_MHC.  
 DR InterPro: IPR001304; Lactin\_C.  
 DR InterPro: IPR000538; Link.  
 DR InterPro: IPR000436; Sushi\_SCR\_CCP.  
 DR Pfam; PF00008; EGF\_2.  
 DR Pfam; PF00047; 1g; 1.  
 DR Pfam; PF00059; Lectin\_C; 1.  
 DR Pfam; PF00084; Sushi; 1.  
 DR Pfam; PF00193; Xlink; 2.  
 DR ProDom; PD000918; Link; 2.  
 DR SMART; SM00032; CCP; 1.  
 DR SMART; SM00034; CLECT; 1.  
 DR SMART; SM00179; EGF\_CA; 1.  
 DR SMART; SM00001; EGF\_Like; 1.  
 DR SMART; SM00409; IG; 1.  
 DR SMART; SM00445; LINK; 2.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE; PS00022; EGF\_1; 2.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01187; EGF\_CA; 1.  
 DR PROSITE; PS01241; LINK\_2; 1.  
 DR PROSITE; PS00615; C-TYPE\_LECTIN\_1; 1.  
 DR PROSITE; PS00441; C-TYPE\_LECTIN\_2; 1.  
 DR GlycoProtex; PS50041; C-TYPE\_LECTIN; 1.  
 KW Glycosite; Proteoglycan; Lactin; Extracellular matrix; Sushi;  
 KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;  
 KW Hyaluronic acid; Alternative splicing.  
 FT SIGNAL 1 26  
 FT CHAIN 27 3562  
 FT POTENTIAL  
 FT VERSION CORE PROTEIN.

DOMAIN 37 136 IG-LIKE V-TYPE DOMAIN.  
 FT 166 243 LINK 1.  
 FT 264 345 LINK 2.  
 FT 3254 3290 EGP-LIKE 1. CALCIUM-BINDING (POTENTIAL).  
 FT 3292 3328 EGP-LIKE 2. CALCIUM-BINDING (POTENTIAL).  
 FT 3341 3435 C-TYPE LECTIN.  
 FT 3460 3518 SUSHI.  
 FT DISULFID 44 129 BY SIMILARITY.  
 FT DISULFID 171 242 BY SIMILARITY.  
 FT DISULFID 195 216 BY SIMILARITY.  
 FT DISULFID 269 344 BY SIMILARITY.  
 FT DISULFID 293 314 BY SIMILARITY.  
 FT DISULFID 3258 3269 BY SIMILARITY.  
 FT DISULFID 3263 3278 BY SIMILARITY.  
 FT DISULFID 3280 3289 BY SIMILARITY.  
 FT DISULFID 3296 3307 BY SIMILARITY.  
 FT DISULFID 3301 3316 BY SIMILARITY.  
 FT DISULFID 3318 3327 BY SIMILARITY.  
 FT DISULFID 3334 3345 BY SIMILARITY.  
 FT DISULFID 3362 3454 BY SIMILARITY.  
 FT DISULFID 3430 3446 BY SIMILARITY.  
 FT DISULFID 3461 3504 BY SIMILARITY.  
 FT DISULFID 3490 3517 BY SIMILARITY.  
 FT CARBOHYD 163 163 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 235 235 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 329 329 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 529 529 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 709 709 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 948 948 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1409 1409 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1479 1479 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1523 1523 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1530 1530 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1625 1625 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1751 1751 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1988 1988 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2088 2088 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2089 2089 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2507 2507 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2642 2642 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2679 2679 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2748 2748 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 2762 2762 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 3069 3069 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 3194 3194 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 3232 3232 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 3545 3545 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT VARSPLIC 485 1411 MISSING (IN ISOFORM V1).  
 SQ SEQUENCE 3562 AA; 388078 MW; 98C566E88C1602D2 CRC64;

Query Match 4.1%; Score 107; DB 1; Length 3562;  
 Best Local Similarity 22.1%; Pred. No. 20;  
 Matches 92; Conservative 46; Mismatches 159; Indels 120; Gaps 22;

QY 47 TLTO---HTTEGANDTILPLNNIQDILVGMKKOKERPVFQVNDATSFYALKTVVP 102  
 DB 984 TWTEGGQISSVTSKESVAALORERQPSVGLPTEKPKFTDV-----TELETVP 1036  
 QY 103 QRITSAAILISDPQOQLAFVNLGFSNTGLQALGIDTDLGDAQFPDQPADANLGGDLS 162  
 DB 1037 QR-----EGDTSLVP---VTVG-----SEDIGEMQVTDHTSFD-----II 1069

QY 163 QWVAPFTTTHGVF---LIGSDODDFLOFTDIOISTFGSS---ITQVALSGSARPPDQ 217  
 DB 1070 HTEAVTSTKASEVPFKELSTKDD-----RELGTAGMSTLPVTSVQM----- 1112

QY 218 AGHEHFGFLDGISQPSVTGME-----TTVFPQGVVPPQIILT-----GRGDGTGT 263  
 DB 1113 --HEQ-KITAGFESPTTTQEKHDMGSAVDENYVATELSVPALMLTEYGVSGVPEVST 1169

QY 264 RPSWALOGSFAFRHQVQPEFNAYTLNAIPA-----NSAGNLTQOEGAEFLGA 314  
 DB 1170 R-SLHLTGTPKAETATDQE-----EKITEAVPVTGTQAKVYESKGTITREED----- 1216

QY 315 RMFGRWKGAPIDLAPTADDP-----LGADP-QRNNNFDYSDTLTDETRCPFGAHR 366  
 DB 1217 RDVGSNWSVLPPHTMLSSPSTAGSISLLTLGLASPSQTPGSGISSELEEVKTVFPSSRAT 1276  
 QY 367 -KTNPRQDLG-----GPVDTFAMRSSIPYG-----PETSDAELASGVTAQDRGL 410  
 DB 1277 DKTTVISDLTSSISAVDKIQTSAKSPFVSSKSPRIPEDEEVTSSDIIVDESI 1333

RESULT 7  
 ID N121\_HUMAN STANDARD; PRT; 1229 AA.  
 AC Q9Y2N3; O75115; Q9Y4S7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Nuclear envelope pore membrane protein POM 121 (Pore membrane protein of 121 kDa) (P145).  
 GN NUP121 OR KIAA0618.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Cordes M., Bauer C., Holmes A.;  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 243-1229 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=98403880; PubMed=9734811;  
 RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H., Nemura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. X. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro."  
 RL DNA Res. 5:169-176 (1998).  
 RN [3]  
 RP SEQUENCE OF 1130-1229 FROM N.A.  
 RC TISSUE=Uterus;  
 RA Koehrer X., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: ESSENTIAL COMPONENT OF THE NUCLEAR PORE COMPLEX. THE REPEAT-CONTAINING DOMAIN MAY BE INVOLVED IN ANCHORING COMPONENTS OF THE PORE COMPLEX TO THE PORE MEMBRANE (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. NUCLEAR PORE MEMBRANE (BY SIMILARITY).  
 CC -!- DOMAIN: CONTAINS F-X-F-G REPEATS.  
 CC -!- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST NUP1, NSP1 AND MAMMALIAN P62 AND NUP153.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
 DR EMBL; AC006014; AAD28064.1; -;  
 DR EMBL; AB014518; BAA31593.1; -;  
 DR EMBL; AL080109; CAB45713.1; -;  
 KW Nuclear protein; Transpore; Transmembrane; Repeat.  
 FT DOMAIN 1 40 CISTERNAL SIDE (POTENTIAL).  
 FT TRANSMEM 41 61 POTENTIAL.  
 FT DOMAIN 62 1229 PORE SIDE (POTENTIAL).  
 FT DOMAIN 4 10 POLY-ALA.  
 FT DOMAIN 51 56 POLY-ALA.  
 FT DOMAIN 294 299 POLY-LYS.  
 FT DOMAIN 441 444 POLY-SER.  
 FT DOMAIN 499 502 POLY-PRO.

```
FT DOMAIN 733 736 POLY-SER.
FT DOMAIN 819 826 POLY-SER.
FT DOMAIN 869 875 POLY-THR.
FT DOMAIN 1061 1067 POLY-SER.
FT CONFLICT 297 297 K -> E (IN REF. 2).
FT CONFLICT 336 336 S -> N (IN REF. 2).
FT CONFLICT 379 379 T -> A (IN REF. 2).
FT CONFLICT 451 451 K -> R (IN REF. 2).
FT CONFLICT 545 545 T -> I (IN REF. 2).
FT CONFLICT 559 559 T -> P (IN REF. 2).
FT CONFLICT 702 702 S -> F (IN REF. 2).
FT CONFLICT 881 881 P -> H (IN REF. 2).
FT CONFLICT 965 965 T -> A (IN REF. 2).
FT CONFLICT 982 984 PAT -> A (IN REF. 2).
FT CONFLICT 991 996 ASTIKI -> PSMIKV (IN REF. 2).
FT CONFLICT 1000 1000 H -> Y (IN REF. 2).
FT CONFLICT 1006 1008 OPT -> Hpl (IN REF. 2).
FT CONFLICT 1044 1044 MISSING (IN REF. 2).
FT CONFLICT 1102 1102 A -> T (IN REF. 2).
FT CONFLICT 1124 1124 T -> A (IN REF. 2).
FT CONFLICT 1146 1146 S -> G (IN REF. 2).
FT CONFLICT 1165 1165 L -> Q (IN REF. 1).
FT CONFLICT 1195 1195 A -> G (IN REF. 1).
FT CONFLICT 1199 1229 SAAPSFISGSGSKTPGARORLOARORHTKK -> NTFARQ
OEHSPPRGNNLSKRLDLAVAAQGPBRGQASSPTTRKE
(IN REF. 2)
P -> L (IN REF. 3)
515655D12858998B CRC64;
SQ CONFLICT 1202 1202
SEQUENCE 1229 AA; 125087 MW; 515655D12858998B CRC64;

Query Match 4.0%; Score 105; DB 1; Length 1229;
Best Local Similarity 22.9%; Pred. No. 6.7;
Matches 83; Conservative 35; Mismatches 153; Indels 92; Gaps 17;

QY 91 TSPKRLAKTYVPQRITSA-----ILISDP--SGOPLAVNLGFSNTGLQALGITD 139
DB 874 TTPSGSLPTAVTATSSADPSGSGTLATSAPTSOP---TLTFSNTSTPTFNTI-- 927

QY 140 DLGDAGPPDGFADANLGDLSQWVAFFTGTTHGVFLI-----GSDDDPFLDQFTDD 193
DB 928 -----PGSSAKSP-----LPSYCANPQPAFGAEGQPPQAKPALTPSFGS- 970

QY 194 ISTRESSITVOALSGSARP-----FDQAGHEHFG-----LD 227
DB 971 -SFTFGNSAAPAPATPAPASTKIYPAHVPTPIQPTGGAHSAFGKATASAFGAR 1029

QY 228 GISQPSVTGMEITVFGQAVVPPGIILTGARDGDTGTRPSMALDGSFMAFHFGQKUPERN 287
DB 1030 ASSQPAF-CGSTAVFSFGAATSSGCGATTGTASSGSSS-----VFESTTP--S 1075

QY 288 ATTL-ANALPANSAG---NLTOQBEAEFLGARMGRKSKGAPIDLAPTA---DDPALGAD 340
DB 1076 PFTFGSAAAPASGSGFGINVATPSSSATTGAFSGAOGSSTATSTPFTGGLCONLGLTT 1135

QY 341 PQRNNFDDYDLTDETRCPFGAHVAKTNPRQDGCQVDFTHAMRSSIPYGPETSDAEIA 400
DB 1136 GQ-STFPANVSTTESKVPFGG---TATPTFGMLTAPAGVGTSGSSLSAGASAPAQGF 1191

QY 401 SGV 403
DB 1192 VGV 1194

RESULT 8
NU42_YEAST STANDARD; PRT; 430 AA.
ID NU42_YEAST
AC P49686;
DT 01-FEB-1996 (rel. 33, Created)
DT 01-FEB-1996 (rel. 33, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Nucleoporin NUP42 (Nuclear pore protein NUP42).
GN NUP42 OR RPL1 OR YDR192C OR YD9346.04C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
```

```
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95360993; PubMed=7634338;
RA Stutz F., Neville M., Rosbash M.;
RT "Identification of a novel nuclear pore-associated protein as a
functional target of the HIV-1 Rev protein in yeast.";
RL Cell 82:495-506(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Oliver K., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -! FUNCTION: INTERACTS SPECIFICALLY WITH THE HIV-1 REV PROTEIN
EFFECTOR DOMAIN AND PROMOTES RNA EXPORT.
CC -! SUBCELLULAR LOCATION: Nuclear pore complex.
CC -! DOMAIN: CONTAINS F-G REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U30614; AA87033.1; -.
DR EMBL; Z48784; CAAB8706.1; -.
DR SGD; S0002600; NUP42.
DR InterPro; IPR004325; Nucleoporin FG.
DR Pfam; PF03093; Nucleoporin FG; 26.
KW Nuclear protein; Transport; Repeat.
FT CONFLICT 331 331 K -> Q (IN REF. 1).
FT CONFLICT 419 419 V -> D (IN REF. 1).
SQ SEQUENCE 430 AA; 42778 MW; 357F949145261F4 CRC64;

Query Match 4.0%; Score 104.5; DB 1; Length 430;
Best Local Similarity 22.4%; Pred. No. 1.8;
Matches 94; Conservative 38; Mismatches 154; Indels 114; Gaps 20;

QY 138 TDDLDGAPDPPGQFADANLGDLSQWVAFFTGTTHGVFLISDQDDPFLDQFTDDISST 197
DB 32 TNNMGSAFGRPSFCTANTM-----TGGTTTSAP-----CMPQFGTNTGNT 72

QY 198 FGSSTIYQVQALSGSARPPDQAGHEHFGFLD--GISQPSVTGMEITVFGQAVVPPGIILT 255
DB 73 GMTSISAFGNSTNAKP--SARGAPAFGSSAPINVPSTT-----SARGAPSF--- 119

QY 256 GRDQDTGTRPSMALDGSFMAFHQOXPPEFNAYTLALIPANSAGNLTOQGAELGAR 315
DB 120 ---GSGT-----FGAM---AATSNPFKSGSGMSAF--GQP 148

QY 316 MGRWKSQAPIDLAPTDADPALGA-----DQRNNFDPYST-- 352
DB 149 ATGANKTALPSSSVSNNSNAFGLAASNTPLTTSPFCSLOQMSQWASSTSSAFGKPTTG 208

QY 353 LIDETRCPPFGAHVAKTNPRQDLG--GPVDTF-----HAMRSSIPYGPETSD 396
DB 209 AATNTQSPPGT-IQNTSTSSGTVSPGTFGTNSNNKSPSNLQSGAGAGASSPFGTTTSK 267

QY 397 AELASGVTRQDRGLLFVEYQSIIGNGFRF---QOINMANNANP-----FSKPTTGC 446
DB 268 ANNNNNVSSAAGCTNNNSPFGSGGCTFQSASLNKNTKNGNQGSSFGKMGKSFGITPON 327

QY 447 E-----PIIGGTTPTTVGGLDP---LNONETFFVPLFVLPKGEYFFLDSISALTAT 495
DB 328 DANKVSQSNPSFGQTMPT---DPNISLSKNGNATSGF-----GQOQNMATNVANATAT 379

RESULT 9
BPRV_BACNO STANDARD; PRT; 603 AA.
ID BPRV_BACNO
```





RA Sevinc M.S., Mate M.J., Switala J., Fita I., Loewen P.C.;  
 RT "Role of the lateral channel in catalase HPII of Escherichia coli.";  
 RL Protein Sci. 6:490-498(1999).  
 RN [6].  
 RP X-RAY CRYSTALLOGRAPHY (1.89 ANGSTROMS).  
 RX MEDLINE=21348730; PubMed=11455600;  
 RA Melik-Adamyani W.R., Bravo J., Carpena X., Switala J.,  
 RA Mate M.J., Fita I., Loewen P.C.;  
 RT "Substrate flow in catalases deduced from the crystal structures of  
 RT active site variants of HPII from Escherichia coli.";  
 RL Proteins 44:270-281(2001).  
 CC -1- FUNCTION: DECOMPOSES HYDROGEN PEROXIDE IN WATER AND OXYGEN, SERVES  
 CC TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE.  
 CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) = O(2) + 2 H(2)O.  
 CC -1- COFACTOR: HEME GROUP.  
 CC -1- SUBUNIT: HOMOTETRAMER.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC -1- INDUCTION: BY ENTRY INTO STATIONARY PHASE.  
 CC -1- SIMILARITY: BELONGS TO THE CATALASE FAMILY. HPII SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M55161; AAA24039.1; -  
 DR EMBL; AE000268; AAC74802.1; -  
 DR EMBL; D90815; BAA20916.1; -  
 DR EMBL; D90816; BAA15513.1; -  
 DR EMBL; D90817; BAA15521.1; -  
 DR PIR; A39129; A39129.  
 DR PDB; 1IPH; 04-SEP-97.  
 DR PDB; 1CF9; 06-APR-99.  
 DR PDB; 1OF7; 26-APR-99.  
 DR PDB; 1GG9; 30-AUG-00.  
 DR PDB; 1GGE; 30-AUG-00.  
 DR PDB; 1GGF; 30-AUG-00.  
 DR PDB; 1GGH; 30-AUG-00.  
 DR PDB; 1GGI; 30-AUG-00.  
 DR PDB; 1GSK; 30-AUG-00.  
 DR Ecogen; Egi0509; Kate.  
 DR InterPro; IPR002226; Catalase.  
 DR Pfam; PF00199; catalase.1.  
 DR PRINTS; PR00067; CATALASE.  
 DR ProDom; PD000510; CATALASE.1.  
 DR PROSITE; PS00437; CATALASE.1.  
 DR PROSITE; PS00438; CATALASE.2.1.  
 KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;  
 KW 3D-structure; Complete proteome.  
 FT ACT SITE 128 128 BY SIMILARITY.  
 FT ACT SITE 201 201 BY SIMILARITY.  
 FT METAL 415 415 IRON (HEME PROXIMAL LIGAND).  
 FT CONFLICT 198 198 L -> F (IN REF. 2).  
 SQ SEQUENCE 753 AA; 84162 MW; 1F034E486A70FB9 CRC64;

Query Match 4.0%; Score 104; DB 1; Length 753;  
 Best Local Similarity 19.9%; Pred. No. 4.1;  
 Matches 123; Conservative 55; Mismatches 205; Indels 234; Gaps 31;

QY 40 PGCAPIPLTQHTTSSG--ANDTILPLNNIGDILVGKKOKERFVFQ-----VNDAT 91  
 DB 38 PAAEPIPPAOGTATAGSLKAPDT-----RNEKLNLEDEYKSGSEVYALTNGVRINDDN 93  
 QY 92 SFF-----TALKTY-----VPORTSA-----ALLI 112  
 DB 94 SLNAGSGPTLLLEDPLREKJTHFDHERIPERIVHANGSAHGYFOYKSLSDITYADPL 153  
 QY 113 SDPSQOPLAFVNL-----GFSNTGLOALGITDDLG-----DA-Q 145  
 DB 154 SDNKTITPVAFVFTVYGAGSADVTVRDIRGFAIKRFTESGIPDVGNNTPIFFIDAIK 213

QY 146 PFDGOFADANLGDLSQWAPARTGTTIHGVFLIGSDQDDFLDQFTDOISSTFGSSITOV 205  
 DB 214 FPD--FVHAVK---BERHWLP--QGOSAR-----DTFWDIVSLQPEFLHVM 254  
 QY 206 QALSSGAPFPDOAGHENGF-----LDGISQPSVTGMEYTFVPGQAVVPPG11LTGRD 258  
 DB 255 WMSMDRGIPRSYRTMEGFIHTFPLINAEKATPVFHHKPLAGKASLYWDEAQKLTGRD 314  
 QY 259 GDTGRPSM-ALD-GSPFARHPQKYPFNATL-----ANAIKANSAGULT 304  
 DB 315 PPFARLEMEALIEADPEYELGFLPDEBERKFPDLDLTKLIPEELVQVQKIV 374  
 QY 305 OOEAGFELGAMFGRWKSQA--PIDLAPTD--DPAAGADPRANNPDSPTLTDETRC 359  
 DB 375 LNRND-----NFAFNEQADHPGHIVGLDFTNDPLL-----QGRLFETYDTQISR--- 422  
 QY 360 PGCAVTRKTRNPDUGGVDTFFAM--RSITY-----GREPSDA----- 397  
 DB 423 -----LGGP--NFHEIPINRPTCEYHNFORDGKMRGIDITNPANYPEPNSIN 466  
 QY 398 -----ELASGVTAODRGHLPVEYOS-----IIGNGFRFOQIMWANN 433  
 DB 467 DWPREPTEPRGPRGSGFESYQERVEGKNVRERSPSFGEYSHPRLSQTREGRHIVDG 526  
 QY 434 ANFPFSKPTTPTGI-EPITGQ-----TPRTVYGL--DP 463  
 DB 527 FSEFLSKVPRPYIRRRVVDLAMIIDLTLAQAVAKNLGIELTDQINITPRPDVNLKKDP 586  
 QY 464 LQNETFTVPLFVJPKG 480  
 DB 587 -----SLSLVAIPDG 596

RESULT 11  
 FOXA\_YEREN STANDARD; PRT; 710 AA.  
 ID FOXA\_YEREN  
 AC 01674;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Ferrioxamine receptor precursor.  
 GN FOXA.  
 OS Yersinia enterocolitica.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Yersinia.  
 OX NCBI\_TaxID=630;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 51872 / WA-C / Serotype O:8;  
 RX MEDLINE=92349959; PubMed=1640832;  
 RA Baemler A.J., Hanke K.;  
 RT "Ferrioxamine uptake in Yersinia enterocolitica: characterization of  
 RT the receptor protein FoxA";  
 RL Mol. Microbiol. 6:1309-1331(1992).  
 RN [2]  
 RP REVISIONS.  
 RC STRAIN=ATCC 51872 / WA-C / Serotype O:8;  
 RA Baemler A.J.;  
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: FERRIOXAMINE BINDING AND UPTAKE, IN ASSOCIATION WITH THE  
 CC TONB PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: Outer membrane.  
 CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X60447; CAA42975.1; -.  
DR PIR; S22673; S22673.  
DR HSSP; P06971; IQJO.  
DR InterPro; IPR000531; TonB\_boxC.  
DR Pfam; PF00593; TonB\_boxC.1.  
DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; 1.  
DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
KW Signal; Receptor; Transmembrane; Outer membrane; Iron transport;  
KW Transport; TonB box.  
FT SIGNAL 1 26  
FT CHAIN 27 710  
FT SITE 28 35  
FT SITE 693 710  
FT TONB C-TERMINAL BOX.  
FT SITE 29 37  
FT TRANSMEM 65 73  
FT TRANSMEM 91 99  
FT TRANSMEM 106 114  
FT TRANSMEM 137 145  
FT TRANSMEM 152 160  
FT TRANSMEM 180 188  
FT TRANSMEM 194 202  
FT TRANSMEM 208 216  
FT TRANSMEM 259 267  
FT TRANSMEM 271 279  
FT TRANSMEM 293 301  
FT TRANSMEM 309 317  
FT TRANSMEM 353 361  
FT TRANSMEM 370 378  
FT TRANSMEM 427 435  
FT TRANSMEM 443 451  
FT TRANSMEM 476 484  
FT TRANSMEM 491 499  
FT TRANSMEM 517 525  
FT TRANSMEM 531 539  
FT TRANSMEM 555 563  
FT TRANSMEM 567 575  
FT TRANSMEM 579 587  
FT TRANSMEM 610 618  
FT TRANSMEM 624 632  
FT TRANSMEM 649 657  
FT TRANSMEM 671 679  
FT TRANSMEM 684 692  
FT TRANSMEM 702 710  
SQ SEQUENCE 710 AA; 78382 MW; 82EB6EC1456900B8 CRC64;  
  
Query Match 4.0%; Score 103.5; DB 1; Length 710;  
Best Local Similarity 21.0%; Pred No. 4.1; Indels 167; Gaps 23;  
Matches 118; Conservative 61; Mismatches 215; Indels 167; Gaps 23;  
  
QY 62 LPLNNIQCD-ILVGMKKOKERFVFQVNDATSPKTKTYVQVITSAAILISDPQSQ--- 117  
DB 20 IPLASIAADTIEVAKAGHEADLPSTGVTATTKGATKDQPLLTQAQSVVTVRQOMDD 79  
  
QY 118 QPLAFVNLGFSNTGLOALGITDDLGAQFDGQFADANLGDLSOWVAPFTGTHGVF 177  
DB 80 QNAVTV-----QALNTPGVFTG-----FSGGATRYDTVA--LRGPHGGVNNVTF 123  
  
QY 178 LIG-----SDDDDF-----LDQTDIDISSTFGSITQVQALSGSARP-PDQAGH 220  
DB 124 LDGLRLSDGGSYNVLQVDWFLERIDVIKGPSSALYGOSIPGVVVMVMTSKRPQTSEGH 183  
  
QY 221 -----EHFGF-LGDISQPSVTGWEITVFGQAVVP----- 249  
DB 184 FLRTAGNNTQVAAPDYTDATSEHAFRLTGTRNSDTNYDHOREERYAIAPSLWQDPE 243  
  
QY 250 -PGIILTG---RDGDTGTRPSWALDGSFM-----AFRHFQOKVPEFNAY 289  
DB 244 NTSLLLRANLQKDPGGVHSAVPADGSIYGOKLSRGFFDGSNNHNVFKRWQO-----IYSY 299  
  
QY 290 TLA-----NAIPANSAGNLTO-QEAGAEFLGARMEGRWKSQ----- 323  
DB 300 EFSHKFDVWSFRQNASYTHNTQLEQVYOGGWNDSORTLMNRYISGEDSSLNAPAVDNLQ 359

QY 324 -APIDLAFTADDPALGADPQNNNNFYSSOTLDTETRCRPGAHVKNTRQDLGGPVDTFH 382  
DB 360 EADLRATAVKKHVKLLGVDFQKFRNLRSDS-----AYATPLNPTVGVSGSTLYS 409  
  
QY 383 AMRSSIPYGPETSDAEL---ASGVTAQDRGLLFVEYQSIIGNGFRFQOINWANNAN--- 435  
DB 410 DYLLTTP-GINTSVLSRRYEOSGVYLODEMTLDNWHLNLSGRYDRMKTENINNTANSTDE 468  
  
QY 436 -----FPFSKPIITPGIEPIIGQITPTPTVTVGGLDPLNQNETHV 472  
DB 469 RTDNHASGRASLLYSFDSIGSPVSVYSOAITPSLFPDAQOKL-----LKPMT-SEQVEV 521  
  
QY 473 PLFVIPKGEYFFLPISALT 493  
DB 522 GIIYQPPGSGTSLYSAALYDLT 542  
  
RESULT 12  
NUPI\_YEAST  
ID NUPI\_YEAST STANDARD; PRT; 1076 AA.  
AC P20676;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Nucleoporin NUPI (Nuclear pore protein NUPI).  
GN NUPI OR YOR098C OR YOR3182C.  
OS Saccharomycetes cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]\_taxid=4932;  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90275616; PubMed=2190694;  
RA Davis L.I., Fink G.R.;  
RT "The NUPI gene encodes an essential component of the yeast nuclear pore complex."  
RL Cell 61:965-978(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97344368; PubMed=9200815;  
RA Voss H., Benes V., Andrade M.A., Valencia A., Rechmann S., Teodoru C., Schwager C., Paces V., Sander C., Ansoorge W.;  
RT "DNA sequencing and analysis of 130 kb from yeast chromosome XV."  
RL Yeast 13:655-672(1997).  
CC -!- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.  
CC NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLLOCATION OF  
CC THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT.  
CC -!- SUBCELLULAR LOCATION: Nuclear pore complex.  
CC -!- DOMAIN: APPEARS TO BE DIVIDED INTO THREE DOMAINS DEFINED BY  
CC OR OF NUP2 IS REQUIRED FOR GROWTH.  
CC -!- DOMAIN: CONTAINS P-X-F-G REPEATS.  
CC -!- SIMILARITY: TO THE CENTRAL REPEATING UNITS OF NSP1 AND NUP2, AND  
CC TO MAMMALIAN P62.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
CC -----  
CC EMBL; M35632; AAA34822.1; -.  
CC EMBL; X94335; CAA64020.1; -.  
CC EMBL; Z75006; CAA99295.1; -.  
CC PIR; A35622; A35622.  
CC SGD; S0005624; NUP1.  
KW Nuclear protein; Transpore; Repeat.  
FT DOMAIN 333 949  
SQ SEQUENCE 1076 AA; 113581 MW; 4AC23567D2PB53CC CRC64;  
  
Query Match 4.0%; Score 103; DB 1; Length 1076;

Best Local Similarity 20.9%; Pred. No. 7.9;  
Matches 89; Conservative 37; Mismatches 182; Indels 119; Gaps 16;

QY 15 GSSSHVNAKLGAROT--RTPLPLTPPGQAPLTLTQHTES---GANDTILPLNNIG 69  
DB 684 GSNTPSFKRPANETDKRPPTSPFTFGSTNNNTTTSTKPSFGADES---MKSTAS 740  
QY 70 DILVGGKKCKEFVFOVNDATSPFKALKTVPQRISAILSDSQOQLAFVNLGFEN 129  
DB 741 TAAANTKLSNFSFTKRNH-----KENSNSPTSPGDSAS 778  
QY 130 TGLQALG-ITDDLGAQPPDG-QPADANLGDLS-----QVAPPT-----G 170  
DB 779 TPIPVGKRTDAGTNTSKSAFSFGTANTGTNASANSFSPFNAPATNGITTTTNTS 838  
QY 171 TTHIGVFLGSDQDDFLDQFDDISSTFGSSITQVALSGSARPPQAGHEHGFIDIS 230  
DB 839 TINIAGTFNVGKRDQSIAGNTNGASAFGS-----SSGTAATGAASNOSFNF----- 887  
QY 231 QPSVTGWEITVPPGQAVVPGIILTRDGDGTGTRPSMALDGSFMAFRHPQKVPENAVT 290  
DB 888 -----GNNAGAGLNPFTSATSTNANAGLFNKXPPSTNAQN 922  
QY 291 LANAIIPANGAGNLTQOEGAEFLGARMFGWKSGAPIDLAPTDADPALGADPOGNNNFYD- 349  
DB 923 VAVPSAFNFTGNGNSTPPGGSVF--NMNG--NTNANTVPAGSNNQPHOSQTPSNTNNSFT 978  
QY 350 -----SDTLIDETRCP--FEAHVRK-----TNPRQDLGG---PVDTF 381  
DB 979 PSTVPINPFGNGIGITVTATNALRPSDIFGANASGSNSNVNPSIFGAGAGVPTTSF 1038  
QY 382 HAMRSS 387  
DB 1039 GQPOSA 1044

RESULT 13  
POL HV1BR STANDARD; PRT; 1015 AA.  
AC P03367;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE POL polypeptide (Contains: Protease (Retropepsin) (EC 3.4.23.16);  
Reverse transcriptase (EC 2.7.7.49); Ribonuclease H (EC 3.1.26.4)).  
GN POL.  
OS Human immunodeficiency virus type 1 (BRU isolate) (HIV-1).  
OC Viruses; Retrovirus; Retroviridae; Lentivirus.  
OX NCBI\_Taxid:1186;  
RN 11)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=8509333; PubMed=2981635;  
RA Wain-Hobson S., Sonigo P., Danos O., Cole S., Allzon M.;  
RT "Nucleotide sequence of the AIDS virus, LAV".  
RL Cell 40:9-17(1985).  
RN 12)  
RP REVISIONS TO 23-35.  
RX MEDLINE=86245056; PubMed=2424612;  
RA Allzon M., Wain-Hobson S., Montagnier L., Sonigo P.;  
RT "Genetic variability of the AIDS virus: nucleotide sequence analysis  
of two isolates from African patients".  
RL Cell 46:63-74(1986).  
RN 13)  
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).  
RX MEDLINE=92190341; PubMed=1799632;  
RA Spinnell S., Liu Q.Z., Alzari P.M., Harel P.H., Poljak R.J.;  
RT "The three-dimensional structure of the aspartyl protease from the  
HIV-1 isolate BRU".  
RL Biochimie 73:1391-1396(1991).  
CC -1- CATALYTIC ACTIVITY: Specific for a P1 residue that is hydrophobic,  
and P1' variable, but often Pro.  
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-  
phosphomonoester.

CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
CC + {DNA} (N).  
CC -1- P1: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
CC DETERMINED.  
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
CC KNOWN AS THE RETROPEPSIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: K02013; -; NOT\_ANNOTATED\_CDS.  
DR PIR: A03966; CMTVLV.  
DR PDB: 1HNP; 15-OCT-92.  
DR MEROPS: A02.001; -;  
DR InterPro: IPR001995; Asparticase\_rtrv.  
DR InterPro: IPR001969; Asparticase\_site.  
DR InterPro: IPR001037; Integrase\_C.  
DR InterPro: IPR003108; Integrase\_Zn.  
DR InterPro: IPR002156; RNaseH.  
DR InterPro: IPR004777; RYase.  
DR InterPro: IPR001584; Rve.  
DR Pfam: PF00075; rnaaseh; 1.  
DR Pfam: PF00077; rvp; 1.  
DR Pfam: PF00078; rvc; 1.  
DR Pfam: PF00552; Integrase; 1.  
DR Pfam: PF00665; rve; 1.  
DR Pfam: PF02022; Integrase\_Zn; 1.  
DR PROSITE: PS00141; ASP\_PROTEASE; 1.  
DR PROSITE: PS00173; ASP\_PROT\_RETROV; 1.  
DR AIDS; Polypeptide; Hydrolyase; Aspartyl protease; Endonuclease;  
KW Nuclease; Transferase; RNA-directed DNA polymerase; 3D-structure.  
FT CHAIN 69 167  
FT ACT SITE 93 93 BY SIMILARITY.  
FT STRAND 78 82  
FT STRAND 87 92  
FT TURN 94 95  
FT STRAND 100 102  
FT TURN 111 117  
FT STRAND 118 119  
FT STRAND 120 134  
FT TURN 135 136  
FT STRAND 137 146  
FT STRAND 152 153  
FT TURN 155 161  
FT HBLIX 162 162  
SQ SEQUENCE 1015 AA; 115031 MW; 164702F074A84394 CRC64;  
Query Match 3.9%; Score 102.5; DB 1; Length 1015;  
Best Local Similarity 22.3%; Pred. No. 8;  
Matches 79; Conservative 47; Mismatches 115; Indels 113; Gaps 20;

QY 78 QKERFV---FPOVNDATSPFKAL--KTYVPQRITSAAILSDSQO--PLAFVNLGFSNT 130  
DB 597 EKEPIVGAFTFVDDAARETKLAKAGVTVNRGRQKVYLTDTNKTQLQAIHLALDQS 656  
QY 131 GLQALGITDD--LGDAPF-PDQPADAN-LGDDLSQ-----VWAPFTGTHGVFL 178  
DB 657 GLEVNIVTDSQVALGIIQAOPDKSESELVNOIITQLIKKEKYLAVPAHKG-----I 709  
QY 179 IGDQDDPDDQDITDSSITFGSSITQVALSGSARPPDQAGHH-----RQF 225  
DB 710 GGNQVQDKL-----VSAGIRKVLFDG---IDKAQDHEKXYSNMWMAASDFVL 755  
QY 226 LDGISOPSVTGWEITVPPGQAV-----VPPGIILTRDGDGTGTRPSMALDGSFMAFRHPQ 280  
DB 756 PIVVAKEIVASCDKQCLGEMAHQGVDCSPGI-----WQDCTHLEGKVL 801  
QY 281 QKVPENAVTLANAIIPANGAGNLTQOEGAEFLGARMFGWKSGAPIDLAPTDADPALGAD 340



RL Yeast 11:61-78(1995).  
 CC -1- FUNCTION: NOT KNOWN.  
 CC -1- PTM: THE N-TERMINUS IS BLOCKED.  
 CC -1- SIMILARITY: CONTAINS 2 EH DOMAINS.  
 CC -1- CAUTION: WAS ORIGINALLY THOUGHT TO BE A SUBUNIT OF PAB-DEPENDENT  
 CC POLY(A)-SPECIFIC RIBONUCLEASE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; 238062; CAB6208.1; -;  
 CC EMBL; M90688; AAA34841.1; -;  
 CC EMBL; X79743; CAB38097.1; -;  
 CC PIR; S30889; S30889.  
 CC PIR; S48440; S48440.  
 CC SGD; S0001445; PAN1.  
 CC InterPro; IPR000261; EPS15\_repeat.  
 CC Pfam; PF00036; ehand; 3.  
 CC SMART; SM00054; EFh; 2.  
 CC SMART; SM00027; EH; 2.  
 CC PROSITE; PS00031; EH; 2.  
 KW Repeat.  
 FT DOMAIN 142 575 15 X 12 AA APPROXIMATE REPEATS.  
 FT REPEAT 142 153 1-1.  
 FT REPEAT 164 175 1-2.  
 FT REPEAT 188 199 1-3.  
 FT REPEAT 215 226 1-4.  
 FT REPEAT 235 246 1-5.  
 FT DOMAIN 270 359 EH 1.  
 FT REPEAT 392 403 1-6.  
 FT REPEAT 409 420 1-7.  
 FT REPEAT 422 433 1-8.  
 FT REPEAT 446 457 1-9.  
 FT REPEAT 467 478 1-10.  
 FT REPEAT 498 509 1-11.  
 FT REPEAT 507 518 1-12.  
 FT REPEAT 538 549 1-13.  
 FT REPEAT 545 556 1-14.  
 FT REPEAT 564 575 1-15.  
 FT DOMAIN 328 680 2 X 23 AA APPROXIMATE REPEATS.  
 FT REPEAT 328 350 2-1.  
 FT DOMAIN 600 689 EH 2.  
 FT REPEAT 658 680 2-2.  
 FT DOMAIN 1084 1125 7 X 6 AA TANDEM REPEATS.  
 FT REPEAT 1084 1089 3-1.  
 FT REPEAT 1090 1095 3-2.  
 FT REPEAT 1096 1101 3-3.  
 FT REPEAT 1102 1107 3-4.  
 FT REPEAT 1108 1113 3-5.  
 FT REPEAT 1114 1119 3-6.  
 FT REPEAT 1120 1125 3-7.  
 FT DOMAIN 1315 1377 8 X 6 AA APPROXIMATE REPEATS.  
 FT REPEAT 1315 1320 4-1.  
 FT REPEAT 1321 1326 4-2.  
 FT REPEAT 1327 1332 4-3.  
 FT REPEAT 1340 1345 4-4.  
 FT REPEAT 1345 1350 4-5.  
 FT REPEAT 1355 1360 4-6.  
 FT REPEAT 1361 1366 4-7.  
 FT REPEAT 1372 1377 4-8.  
 FT DOMAIN 13 22 POLY-GLN.  
 FT DOMAIN 29 34 POLY-GLN.  
 FT DOMAIN 98 106 POLY-GLN.  
 FT DOMAIN 1400 1406 POLY-PRO.  
 FT DOMAIN 1452 1455 POLY-GLU.  
 FT DOMAIN 1474 1480 POLY-PRO.  
 FT CONFLICT 235 235 P -> T (IN REF. 1).  
 FT CONFLICT 266 273 ITADQAK -> YYCPRSGKN (IN REF. 1).

FT CONFLICT 474 487 MISSING (IN REF. 1).  
 FT CONFLICT 653 657 MISSING (IN REF. 1).  
 FT CONFLICT 1291 1291 A -> R (IN REF. 1).  
 FT CONFLICT 1396 1480  
 FT  
 FT  
 FT  
 FT  
 FT  
 SQ SEQUENCE 1480 AA; 160267 MW; F518495FF759553 CRC64;  
 Query Match 3.94; Score 101.5; DB 1; Length 1480;  
 Best Local Similarity 21.84; Pred. No. 16;  
 Matches 100; Conservative 54; Mismatches 176; Indels 129; Gaps 24;  
 QY 34 PLLTNF-----PGQALPTLTQHTT-----ESGANDTILPLNNIQDILVGMKKOKERFVFQ 86  
 DB 218 PQGTGFVLQPNQNAFLBKLKPTAGFVNSFAN-----NGLNNDI-----KIPALISFTT 267  
 QY 87 VNDATSFALKTYVQRIITSAAIILDSQPLAVNIGFSNTGLQ--ALGITDDLGD 143  
 DB 268 AODQAKFETLFRSIV---TNGSNTVSGANGCRKI-LMRSGLPSSQLARIWTLCTDSKAGE 322  
 QY 144 AOPDGPADAAALGDLSQWVAFPTGTTIHGVFLIGSDQDDFLDQFTDIDISTGSSIT 203  
 DB 323 LTFP--EFALAMHLINDVLO-----GDTI--PYELDSKTKREVSSFTDAINLSTANG-- 370  
 QY 204 OVALSGSARPPDO---AGHEHFGFLDGISQPSVTGMEITVPGQAVVPRGIIILGRDGD 260  
 DB 371 DSSANDAPKTPEDERTITGVONT-----QPOPTGMPQTSFG--IFLOSQITGGGVA 420  
 QY 261 TGRPSWALDGSFMAFRFQKVDFNAVYTLANAIPANSAGNLTQOEGAEFLGAMFGRW 320  
 DB 421 SALNPQ---STGFMAPTTF-----INSMNTGTGRLNPQITG-- 453  
 QY 321 KSGAPIDLAFTADDPAL-----GADPQRNNNFYSDTLIDFERCPGKAVRKTNRDGLG 375  
 DB 454 --GAPASQNPITGNALOPQTTGMPPTTGMMPOTTGMMFOT--SFGVNL----- 499  
 QY 376 GVDTFHAMRSSIPYGPETSDAELASGVTAQDRGLLFVEYQSIIIGNGFRFQOINNNAN 435  
 DB 500 -----GPELTGALQSYTG-----GYGSVMP-----QSGSPASMPN 531  
 QY 436 FPPSKPTTPGIEPIITIGOTTPTVSGLDPLNONTFTVPL 474  
 DB 532 LSFNQ---OGLQSOLTGLOPQPTGFLRP--SNFSATMPL 565  
 Search completed: November 27, 2002, 13:46:02  
 Job time : 24 secs

**THIS PAGE BLANK (USPTO)**



```
Db 121 AFVNLGFSNTGLQALIGITDLDGDAQFPDQGFADANLGGDLGQWVAPFTGTTIHGVFLIG 180
Qy 181 SDQDDFLQDFTDDISSTFGSSITQVQALSGSARPFQAGHEHFGFDGIGSQSVTGMWTT 240
Db 181 SDQDDFLQDFTDDISSTFGSSITQVQALSGSARPFQAGHEHFGFDGIGSQSVTGMWTT 240
Qy 241 VPPGQAVVPPGIIITGRDGTGRPSWALDGSFMAFRHQKQVPEFNAYTLANAI PANSA 300
Db 241 VPPGQAVVPPGIIITGRDGTGRPSWALDGSFMAFRHQKQVPEFNAYTLANAI PANSA 300
Qy 301 GNLTOCEGAELFGARMGRKWSGAPIDLAPTADDPALGADPQNNNFYSDTLTDETRCP 360
Db 301 GNLTOCEGAELFGARMGRKWSGAPIDLAPTADDPALGADPQNNNFYSDTLTDETRCP 360
Qy 361 FGAHVKNTPRQDLGGPVDTFHAMRSSIPYGPETSDAELASGVTAQDRGLLFEVQSIIIG 420
Db 361 FGAHVKNTPRQDLGGPVDTFHAMRSSIPYGPETSDAELASGVTAQDRGLLFEVQSIIIG 420
Qy 421 NGFRFQOINWANNANFPFSKPIITPGIEPIIGQTTPTVGGDLPLNQNETFTVPLVPIK 480
Db 421 NGFRFQOINWANNANFPFSKPIITPGIEPIIGQTTPTVGGDLPLNQNETFTVPLVPIK 480
Qy 481 GGYFFLPSISALTATIAA 498
Db 481 GGYFFLPSISALTATIAA 498

RESULT 2
P87212 PRELIMINARY; PRT; 499 AA.
ID P87212;
AC P87212;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Peroxidase.
GN CP021.
OS Polyporaceae sp.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Aphyllophorales; Polyporaceae.
OX NCBI_TaxID=54443;
RN [1]
RP SEQUENCE FROM N.A.
RA Han Y.-H.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U70703; AAB54908.1; -
DR IncerPro; IPK000504; RNA_rec_mof.
DR PROSITE; PS00030; RRM_RNF_1; UNKNOWN 1.
SQ SEQUENCE 499 AA; 53973 MW; EFBDA18EB6B88624 CRC64;

Query Match 53.7%; Score 1396.5; DB 3; Length 499;
Best Local Similarity 55.3%; Pred. No. 4.8e-96;
Matches 283; Conservative 63; Mismatches 145; Indels 15; Gaps 9;

Qy 1 MDLSLFVSVAVILVGSSSHVNAKLGARQTRTPTLLNTPGQAPLPTLTOHTTESGANDT 60
Db 1 MOLKHLAATAAF--SAVTSFAFYHVKRARSSPLIGSPGQPPPLPTAEVQSSAGNDS 58
Qy 61 ILPLNINIGDILVGMKKQKRFVFFQVNDATSEKTKALTYVVPQRTISAAILSDPSQOPL 120
Db 59 -LPFENIGDILVGMKKQKRFVFFHINNATTFKSLKYAPANITSVQTIIGPASGQPQ 117
Qy 121 AFVNLGFSNTGLQALIGITDLDGDAQFPDQGFADANLGGDLGQWVAPFTGTTIHGVFLIG 180
Db 118 AFVNLAFSHTGFGALGVADDDLDQTAFTAGQFADAPFGDDTSTWEAFKGNVDGVFLIG 177
Qy 181 SDQDDFLQDFTDDISSTFGSSITQVQALSGSARPFQAGHEHFGFDGIGSQSVTGMWTT 240
Db 178 SDDVTTTQYRDLKAKLGDAWTVLLDLSAARFGAEKGHEHFGFDGIGSNPTIPGFCPT 237
Qy 241 VPPGQAVVPPGIIITGRDGTGRPSWALDGSFMAFRHQKQVPEFNAYTLANAI PANSA 299
Db 238 -FPGQVVDGVSITFAGRTNPDVTRNPSWALDGSFLVFRKLQVLVPEFFHAKLNDNAL-QNQ 295
```

```
Qy 300 AGNLTQOGEAEFLGARMGRKWSGAPIDLAPTADDPALGADPQNNNFY-----SDTLTD 355
Db 296 AGNLTVEGALLGSRMFGRWNSGAPIDLTTPVDPTLGNDRQNNDFNYIHPGEBLTTD 355
Qy 356 ETRCPFGAHVKNTPRODL--GGPVDTPHAMRSSIPYGPETSDAELASGVTAQDRGLLF 412
Db 356 ETRCPFTAHVKNTPR--DLEAQGLIPDLFAIRAGTTPYGPETVDAESNNTTSDRGLAF 414
Qy 413 VEYQSIIGNFRFQOINWANNANFPFSKPIITPGIEPIIGQTTPTVGGDLPLNQNETFTV 472
Db 415 VEYQSVISNGFRFQOINWANNANFPFNKSEPLGLDVPVIGQGTQGT-FGLDPRNASDSLTI 473
Qy 473 PLFVTPKGGVEYFFLPSISALTATIAA 498
Db 474 POIISNGGEYFFSPSITALVEHFGA 499

RESULT 3
Q9RZ08 PRELIMINARY; PRT; 460 AA.
ID Q9RZ08;
AC Q9RZ08;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Peroxidase, putative.
GN DRA0145.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1259;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Uterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL: AE001862; AAP12218.1; -
DR TIGR; DRA0145; -
KW Peroxidase; Complete proteome.
SQ SEQUENCE 460 AA; 51657 MW; CB8C476C1DDIAB50 CRC64;

Query Match 15.9%; Score 412.5; DB 16; Length 460;
Best Local Similarity 29.3%; Pred. No. 1.3e-22;
Matches 148; Conservative 68; Mismatches 190; Indels 99; Gaps 25;

Qy 34 PLLTNFPQAPLPTLTOHTTESGANDTI-LPLNNTQGDILVGMKKQKER-FVFFQVNDAT 91
Db 2 PEMTLF---KKLRELNVH-----NDKIDLDLDIDQATVL-----RERPEYTGTHAV 46
Qy 92 SFKTA-----LKYTVPQRTISAAILSDPSQOPLAFVNLGFSNTGLQALGTTDGLDAQ 145
Db 47 RFDTAEGGRELLKRLLP-HIASA-----EKWVDVKYAWTAAATSYEGKLGVQDSDLS- 100
Qy 146 FPD-----GQFADAANLGD-----DLSQWVAPFTGTTIHGVFLIGSQDDO-----FLDQFTDD 193
Db 101 PFESFKVMAGRAEHLFDVGENDPFKHEKFPFGTQGVHQLALTIFAENBENWQALVIAHE 160
Qy 194 ISSTFGSSITQVQALSGSARPFQAGHEHFGFDGIGSQSVTGMWTTVPQQA-VVPPGI 252
Db 161 LEATKG--VTLMLREDFGAQP---DSRNSLGYKDMISNPAIEGSGIKFPFGQPAIKPGE 215
Qy 253 ILTGEDDGTGT-----RPS-WALDGSFMAFRHQKQVPEFNAYTLANAI PANSA 306
Db 216 FVLGPGGAGVPLGNPKPEVLGKNKGTFAVRKTHTNAGSFNRYLKENA--EYTGDD----- 269
```



```

QY 307 EGAEFLGARMFGKSGAPIDLAPTADPALGADPORNNNFQYSDTLTDETRCPGAHVR 366
DB 270 --BELLAAKLVGRMSSGAPLTLAPKEDDEPELGHPNNNNFTYKND--PEBLEVPLGSHR 326
QY 367 KTNPRO---DLGCPVDTEPHAMRSSIPYGPETSDAELASGVTAODRGLLFEVQSIIGNGF 423
DB 327 RMNPRDTKLELTVDNIRIRIRATAYGPAYDPKADSLADKVERGLYFISAKAMDTT 386
QY 424 RPOQJNMANNANFPFSKRTITPGIEPIIGOTTPTRTVGGLDPL---NONEFTVTP----- 473
DB 387 EFLQKEMINKAF-----IGQSSER-----DPLVGLQDEDLFTLPKEPVRO 428
QY 474 -----LFPVPGGEYFPPLPSISAL 492
DB 429 RLRGMDTFNVLRGCEYLFMPISAL 453

RESULT 4
OBYMMO PRELIMINARY; PRT; 469 AA.
AC OBYMMO;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical protein Alr1585.
GN Alr1585.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kunitz T., Sasamoto S.,
RA Matsumoto A., Iriuchihara N., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimizu S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Ref. 8:205-213(2001).
DR EMBL; AP003586; BAB79951.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 469 AA; 53500 MW; 6532B596C3C8D4C1 CRC64;

Query Match 14.9%; Score 386.5; DB 16; Length 469;
Best Local Similarity 29.9%; Pred. No. 1.2e-20;
Matches 144; Conservative 57; Mismatches 185; Indels 95; Gaps 23;

QY 64 LNNIGODILVGMKKQKRFV--FOYNDATSEFKTALKTYPQRTSA-----AIIISDP 115
DB 28 LNDLQGNILKGGHSHVHLEQFKPEQVEVVKQMIQSFQYTSYKKQADDAFPRYRK 87
QY 116 SQOPLAFVNLGFSNTGLQALGIT--DLDGAOFDPGQFAD--AAALGD--DLSQWVAPFG 170
DB 88 GVGSDVFANFLSRGVEYLEIEPQIDKPRFMGMKNEIRSLDPRKATWELGFGQ 146
QY 171 TTHGVLFGSQODDFLOFTDDISTRGSSITQVQALSGARF---DQAGH--BHPRF 225
DB 147 SEHIALVLLADDDIVDLQIVQIT---OKLROIATIVHREDGFILRNAGGIIHFPG 202
QY 226 LDGISQPSVTGWEVTFPGQAVV-----PPGIIILTRGDGTGRSMAL 269
DB 203 VDGVSQP-----LFMKRDVVRERVNANCFDKWDPKALPDSITLV--EDNGTKDSY-- 251
QY 270 DSGFMAFRHQKQVPEFNAYTLANAIPNSAGNLQDGAFFLGARMFGKSGAPIDLA 329
DB 252 --GSYLVYRKLEQVNAKAFREDQKLLQKINIGENLA-----GALTGKRFADGTFTL- 301
QY 330 PTADDPALGADPORNNNFQYSDTLTDETRCPGAHVRKTNPRODLAGPVDT----- 380
DB 302 --SDIFTYAVTP--TNNFVYDGDILA--ATKCPHSHTRKTNPRGDTARLLLTDDGHFDEAPK 356
QY 381 ----PHAMRSSIPYGPETSDAELASGVTAODRGLLFEVQSIIGNGFPOJNMANNANFP 436

```

```

DB 357 EERGRHITRAVSYENNPSEKPVSG-----SGILFLCFQSNLENQFNMQSRMANPQNF 411
QY 437 PFSKPIITPGIEPIIGOTT-----PRTVGGLDPLQNNETFTVPLFVLPKGEYFPPLPSISA 491
DB 412 ---VOYNTGPDDLQIQPSTGOKWPKWG--EP--ETEEVNFQLMINKGGEVFAPASISF 464
QY 492 L 492
DB 465 L 465

RESULT 5
OBYMMO PRELIMINARY; PRT; 421 AA.
AC OBYMMO;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Possible periplasmic protein.
GN MA0282.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=2022556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holtroyd S.,
RA Jagsels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrett B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491."
RL Nature 404:502-506(2000).
DR EMBL; AL162752; CAB83589.1;
KW Complete proteome.
SQ SEQUENCE 421 AA; 46119 MW; A598BF5DF4B2F562 CRC64;

Query Match 7.1%; Score 185; DB 16; Length 421;
Best Local Similarity 23.2%; Pred. No. 1.1e-05;
Matches 88; Conservative 41; Mismatches 125; Indels 126; Gaps 17;

QY 135 LGITDLDGAOPDPGQFA-----DAAN-----LGDPLSQWVAPFGTTIH 174
DB 130 LITTVGVSSSLF--DGRFGAKDKKPIHLEMRDFFENDKLQKSWDGDLSLQICAFPTETQ 188
QY 175 GVFLIGSDODDPLDFTDDISTRGSSITQ--VOALSGARPPDQAGHFGFLDGISOP 232
DB 189 AA-----LRDIIKRTVGTAVIRWSIDGWQKSEPGAMAARNLLGFRDGTGNP 235
QY 233 SVTGWETTVFPGQAVVPGIILTRGDGTGRFPMALDGSFMAFRHQKQVPEFNAYTLA 292
DB 236 KVSDEPKTA-----DEVLWTVGAANSLSDEPFMAKXGYSQAVRLIRHVEWMDTPL- 285
QY 293 NAIPNSAGNLQDGAFFLGARMFGKSGAPIDLAPTADDPALGADPORNNNFQYSDT 352
DB 286 -----OEOQTIDFGRRKY-----SGAPMDGKKEADODPFAKDPKPGNTT----- 322
QY 353 LITDETRCPGAHVRKTNPRODLAGP--VDTEPHAMRSSIPYGPETSDAELASGVTAODRGL 410
DB 323 -----PKDSHILRLANPRD-----PEFLAKKRLRLRRAYSV-----SGLGASSG--QLDVL 365
QY 411 LFEVQSIIGNGFPRFOQJNMANNANFPFSKRTITPGIEPIIGOTTPTRTVGGLDPLQNNET 470
DB 366 VFVQYQANLADGFIFVQ--NLNG-----EPLLE----- 392
QY 471 TVPLFVLPKGEYFPF--LPSI 489
DB 393 ----YISFGGCGYFFVPLGV 408

```

```

RESULT 6
ID O9K1P5 PRELIMINARY; PRT; 421 AA.
AC O9K1P5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein NMB0036.
GN NMB0036.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson M.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
RA Cotton M.D., Uitterlinden T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masiognani V., Pizzi M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RL Science 287:1809-1815(2000).
DR EMBL; AE002362; AAP40507.1; -.
DR TIGR; NMB0036; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 421 AA; 46118 MW; 148BF54783D74337 CRC64;

Query Match 6.8%; Score 177; DB 16; Length 421;
Best Local Similarity 23.4%; Pred. No. 4.5e-05;
Matches 89; Conservative 41; Mismatches 124; Indels 126; Gaps 18;

OY 135 LGITDLDGADPPGQFA-----DAAN-----LGDLSQWVAPFTGTH 174
DB 130 LTVTVGSSLF-DGRFLGDKKPIHQEMRDFSNKLOKSCWGDLSQICAFPTQCQ 188
OY 175 GVFLIGSDQDFLOQFTDDISSTGSSITQ--VQALSGSARPPDQAGHEHFGFLG 232
DB 189 AA-----LRDIKHTVQTAVIRMSIDGQPKSEPGAMAARNLLGFRDGTGNP 235
OY 233 SVTGWETTTFVFGQAVVPPGCIILTRGDGTGTRPSWALDGSFPAFRHQKVPFNAYTLA 292
DB 236 KVSDDPKTA-----DEVLTWGAANSIDPEWAKNGSYQAVRLIRHFVEFWDTPL- 285
OY 293 NAIPANSAGNLTOQEGAEFLGARFGRKSGAPIDLAPTDALDGPQNNNFDYSDT 352
DB 286 -----QEQTDIFGRKRY-----SGAPMDGKKEADQDPFAKDP----- 318
OY 353 LTDETRCFGHAHVKTNPQDLGDP--VDTFHAMRSSIPYGPTSDAELASGVTADRG 410
DB 319 -GDIT--PKDSHIRLANPRD-----PEFLKXHLRFRAYS-----SGLASSG--QLDVGL 365
OY 411 LRVEYQSIIGNGFRFQOINMANNANFPFSPITPGIEPIIGTTPRTVGGLDPLNQNTF 470
DB 366 VFVCVQANLADGFIQVQ-NLLNG-----EPLSE----- 392
OY 471 TVPLFVFPKGEYFF-LPSI 489
DB 393 ----YISPGGGYFVPLPGV 408

RESULT 7
ID O8ZF62 PRELIMINARY; PRT; 434 AA.
AC O8ZF62;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

DE Hypothetical protein YPO1856.
GN YPO1856.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkehill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Pringle M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
DR EMBL; AJ414150; CAC90673.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 434 AA; 47800 MW; AA93E8F8623D41B6 CRC64;

Query Match 5.5%; Score 143.5; DB 16; Length 434;
Best Local Similarity 22.9%; Pred. No. 0.015;
Matches 81; Conservative 38; Mismatches 123; Indels 111; Gaps 16;

OY 145 QPPDQGFADAAANLGDLSQWVAPFTGTHGVFLIGSDQDQDFLOQFTDDISSTF--GSS 201
DB 173 RFPNDSLDAGLCHGDVWLQICANTNEVIAL-----RDIIKHTPDLSSVRKKEGFI 225
OY 202 ITVOALSGSARPPDQAGHEHFGFLGDISQPSVTGWETTTFVFGQAVVPPGCIILTRGDGT 261
DB 226 SAHAARSXGQDTPINL-----LGFKQGTANPKIS-----NKLINNVVWSNNAE- 271
OY 262 GTRPSWALDGSFPAFRHQKVPFNAYTLANAI PANSAGNLTOQEGAEFLGARFGRWK 321
DB 272 ---PANAVGGSYQVVRIRFKV-EFWDRT-----PQEQ-----TIFGRDK 309
OY 322 -SGAPIDLAPTDALDGPQNNNFDYSOTLTDTRCPGHAHVKTNPQDLGSGPVD 380
DB 310 NSGAPLQNGHEHDEPNKADPEK-----VIPMDAHLRLANP-----TTET 351
OY 381 FH--AMRSSIPYGPTSDAELASGVTADRGLLFVYQSIIGNGFRFQOINMANNANFPF 438
DB 352 QRNLMLRGYSYSGVSN-----SG--QLDMGLLFVCYQSDLAQAFLTQERLNGALBEY 405
OY 439 SKPIPTGIEPIIGTTPRTVGGLDPLNQNTFTVPLFVFPKGEYFF-LPSIS 490
DB 406 VKPI-----GGGYFTPLPGVA 421

RESULT 8
ID O8XAS4 PRELIMINARY; PRT; 423 AA.
AC O8XAS4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Orf, hypothetical protein.
GN YCDB_012521 OR ECS1265.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
```

RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.:  
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7."  
 RL Nature 409:529-533(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakiwa C., Ogasawara N., Yasunaga T.,  
 RA Kubura S., Shiba T., Hattori M., Shinagawa H.:  
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*  
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 DR EMBL: AE005301; AAC55637.1; -  
 KW EMBL: AP002554; BAB34688.1; -  
 KM Complete proteome.  
 SO SEQUENCE 423 AA; 46620 MW; 25FB987670B750C CRC64;

Query Match 5.1%; Score 133.5; DB 16; Length 423;  
 Best Local Similarity 22.3%; Pred. No. 0.081;  
 Matches 81; Conservative 29; Mismatches 119; Indels 135; Gaps 14;

QY 145 QPDDGQFADANLGDLSQWVAPFTGTHGVFLISDQDDFL-----DQFTDDISSTF 198  
 DB 162 REPDSLDALCHGVLDICANTDVIHALRDIYHTPDLSSVRKKEGFLSDHAAS 221  
 QY 199 GSSITQVALSGSARPFDOAGHEHFGFLDGISQPSVTG-----WETTFPGQAVVP 250  
 DB 222 KKEFTPINL-----GFKDGTANPDGNDKLMQKVWVTA----- 256  
 QY 251 GIIITGRDQDGTFRPSMALDGSFMAFRHQKQPEPNAYTLANIPANSAGNLTOEGAE 310  
 DB 257 -----DOGEFANITIGSYQAVRLOFRV-EFMDRT-----ELKEQD--- 291  
 QY 311 FLCARFEGMK-SCAPIDLAFTADDPALGADPQRNNPDYSDTLTDETRCFGAHVAKTN 369  
 DB 292 ----TIFGDKOTGAPLGMQHEHDVDPYASDPEKGG-----TALDSHRLAN 334  
 QY 370 PRODLGCPVDTFHAKRSSIPYGPETSDAELASGVTAQ--DRGLFVEYOSIIGNGFRQ 426  
 DB 335 PRT---AESSESLMRGVSY-----SLGTTNSGOLDMLGFLFYCHDLEKFFLV 382  
 QY 427 QINMANNANPFSKPTPGIEPIIGQTTPTRVGGDLPLNQNFTVPLFVIRKGEYFF- 485  
 DB 383 OKRLNGEALLEVYKPI-----GGGYFFA 405  
 QY 486 LPST 489  
 DB 406 LPGV 409

RESULT 9  
 Q8Y366 PRELIMINARY; PRT; 1672 AA.  
 AC 08Y366;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Putative hemagglutinin-related protein.  
 GN RSC0115 OR RS00985.  
 OS *Ralstonia solanacearum* (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 OC Ralstonia.  
 NC NCBI\_TaxID=305.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GM1100;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Arlat M., Billault A., Brotier P., Camus J.C., Catolico L.,  
 RA Chandler M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,

RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,  
 RA Stigter P., Theault P., Whalen M., Wincker P., Levy M.,  
 RA Weisenbach J., Boucher C.A.:  
 RT "Genome sequence of the plant pathogen *Ralstonia solanacearum*."  
 RL Nature 415:497-502(2002).  
 DR EMBL: AL646057; CAD13643.1; -  
 KW Complete proteome.  
 SO SEQUENCE 1672 AA; 164046 MW; 55058448B10FE1B8 CRC64;

Query Match 4.9%; Score 128; DB 16; Length 1672;  
 Best Local Similarity 21.3%; Pred. No. 1.7; Indels 106; Gaps 23;  
 Matches 105; Conservative 67; Mismatches 106; Indels 106; Gaps 23;

QY 30 TRTPPLTNFPGQAPLPTLTOTHTESGANDTILPLNNIQGDLVGMKKKQKERVFPQVND 89  
 DB 701 TRYSLTVAAPTALPTTLGATVSSA-----YSGSV 733  
 QY 90 ATSFKALTYPQRTSAAILISDPSQPLAFNLGFSNTGLGALITDDGDAFPQG 149  
 DB 734 AASGTAPTYVA--VTGAL-----GLSLSSAGV-----LSGPTAG 771  
 QY 150 QFADANLGDLSQWVAPFTGTHGVFLISDQDDFLDQFTDDISSTGSSITQVALS 209  
 DB 772 SFEVITIS-ATDSTTSGPFTGSRAV-TLVGSPPLTISPASTAGLTMAGTSVSGSFSAG 829  
 QY 210 GSARPFDOAGHEHFGFLD-GIS---QPSYGMETTFPGQAVVPQIT--LTGRDGT 261  
 DB 830 GGVSPYTAALTVNTGTMPAGLSFMAASATLSGPTTA-----GTVSFTVATDSS 880  
 QY 262 GTRPSMALDGSFMAFRHQKQPEPNAYTLANIPANSAGNLTOEG-AEFLGARMF 317  
 DB 881 GAGP-YAVSGYV-----LTVASPLTLVAPATLPALGTAQSITASSGTATYAVTS 935  
 QY 318 GRKSGAPIDLAFTADDPALGADPQRNNPDYSDTLTDE-----RCPRGAHVAKTN 369  
 DB 936 GALPAGLSLSA-----GLVSGTPAGSGFETVATDANSFTASRAYSITIGATVALN 990  
 QY 370 PRODLGCPVDT-----TFHAMRSSIPYGPETSDAELASGVTAQD-RGLFVEYOSIIGNGF 423  
 DB 991 PATVPATLNTAYSGQTFASGGIGFYTAVASGTLPAVGSLNSTGVLGSLGTPALGSSSTF 1050  
 QY 424 RFOQINMANNANPFSKPTPGIEPIIGQTTPTRVGGDLPLNQNFTVPLFVIRP---K 479  
 DB 1051 STRATDSTGAGAPYTG--TRGYTLVWGQ---AIGTAPVTAATTSTVPLVPLHPTANAT 1104  
 QY 480 GCEYFFLPSTSA 491  
 DB 1105 GGPFSSTVIVAA 1116

RESULT 10  
 Q9RZ03 PRELIMINARY; PRT; 1067 AA.  
 AC 09RZ03;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Extracellular nuclease, putative.  
 GN DRB0067.  
 OS *Deinococcus radiodurans*.  
 OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;  
 OC Deinococaceae; Deinococcus.  
 NC NCBI\_TaxID=1299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RI;  
 RX MEDLINE=20036896; PubMed=10567266;  
 RA White O., Eissen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crobby M., Shen M.,  
 RA Vamathevan J.J., Lam P., McDonald L., Uricarback T., Zaleski C.,  
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,

RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
RA Fraser C.M.;  
RT "Genome sequence of the radioresistant bacterium Deinococcus  
RT radiodurans R1";  
RL Science 286:1571-1577(1999).  
DR EMBL; AE001826; AAF12592.1; --  
DR TIGR; DRB0067; --  
DR InterPro; IPR005135; Exo\_endo\_phos.  
DR Pfam; PF03372; Exo\_endo\_phos; 1.  
KW Plasmid; Complete proteome.  
SQ SEQUENCE 1067 AA; 109780 MW; D4D3B79B38F72B30 CRC64;

Query Match 4.9%; Score 127.5; DB 16; Length 1067;  
Best Local Similarity 23.0%; Pred. No. 0.93;  
Matches 121; Conservative 60; Mismatches 171; Indels 175; Gaps 28;

QY 13 LVSSSHYNAA---KLGARQTRTTP-----LTPPGQAP 44  
DB 55 LSGAGSAIKQASVQLGAQTEVGLGSKNNYGVVSSNNTLRWATFVNTSDQ-P 113  
QY 45 L---PTLTQHTTESGANTPI--LPLNNIQDILVGMKKQKRFVFFQVNDATSF----- 93  
DB 114 LAAPTPIPVDTGAGGTIGTAPKNR-----YFGSDASSRABLLAPD 157  
QY 94 -----KTALKTVVQRTISAAILSDPSOQPLAFVNLGFSNTG--LQAL--GITDDLGA 144  
DB 158 SSNDPKTAVKTLRLDLDGSGVQVNLPAQLQA---GTSHTGWMKALPAGAT----- 207  
QY 145 QPDPQFADANLGGDLSQWVAPRTGTTIHGVFLGSDQDDFLDQFTDIDISSTFSSITQ 204  
DB 208 ---GQVTLASIFAASNSQNPFS-----FDLVFTVADVNPSTTLTNIGA 249  
QY 205 VQALS-GSARFPDQAGHEHFGFLDGSQPSVTGWTETTFPGQAVVPPGIIITG----RDG 259  
DB 250 VQGTTPGGRPAALTGAQ-----TIEGVVTAVEPG-----LSGFFVQEEG 289  
QY 260 DTGTFPSWALDGSFMAHRHFOQKYPEFNATLANAIPANSAGNLQOEG--AEFLGARMF 317  
DB 290 IDADRETTSDGLF---PPAAELKLPDKTQERYEGMRVRFPETLITNNYTYGRY--- 388  
QY 318 -----GRWKSGAPIDLAPTADDPALGADPQNNNF-----YSDTLTDTETCPFGAHVRK 367  
DB 335 TAPTPTKLLSLGLA---PPAAELKLPDKTQERYEGMRVRFPETLITNNYTYGRY--- 388  
QY 368 TNPRDLGSPVDTTHMRSSSPYGPETSDAE-----LASGVTAQ-----DRGLLFVYQSI 419  
DB 389 -----GQLDLSNAGRMFNFTNGNASASEQSTIILDDGVSAQNPLDNLVLSABTRRT 440  
QY 420 GNGFRPQINMANNANFFSKPIPTPGIEPIIGQTT-----PRTVGG 460  
DB 441 GDTVTGLSGVHVSANQPMLEP--EGAVEFVSANSRASNALPHDVGG 485

RESULT 11  
Q8ZCG6 PRELIMINARY; PRT; 299 AA.  
AC Q8ZCG6;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DE Hypothetical protein YP03025.  
GN YP03025.  
OS Yersinia pestis.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Yersinia.  
OX NCBI\_TaxID=632;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;  
RX MEDLINE=21470413; PubMed=11586360;  
RA Parkhill J., Wren B.W., Thomson N.R., Titchell R.W., Holden M.T.G.,  
RA Prentice M.B., Sebaiha M., James K.D., Churcher C., Mungall K.L.,  
RA Baker S., Baeham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,

RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
RA Letwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,  
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,  
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;  
RT "Genome sequence of Yersinia pestis, the causative agent of plague";  
RL Nature 413:523-527(2001).  
DR EMBL; AJ414155; CAC92267.1; --  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 299 AA; 33389 MW; 5569F04D38F8DBC3 CRC64;

Query Match 4.9%; Score 126.5; DB 16; Length 299;  
Best Local Similarity 21.7%; Pred. No. 0.16;  
Matches 91; Conservative 45; Mismatches 109; Indels 175; Gaps 23;

QY 109 AILISDPSOQPLAFVNLGFSN--TGLQALGITDDLGAQFPD-----GQFADANLGGDLS 162  
DB 16 AIFMEAKVQGGDLAIRLGCKKFCQSLQEL-----QQQFPDDEHLGAVIAFGSNVMDLS 68  
QY 163 -----OWVAPPTGTTIHGVFLIGSDQDDFL---DQFTDDISSTFGSSITQVQALSGSARP 214  
DB 69 NGQAGAKELKPVPL---GKGLAPATQDRLJIHQSLRQDINFTLAQAV---AAGSAIA 122  
QY 215 FDQAGH-----EHFGFLDGSQPSVTGWTETTFPGQAVVPPGIIITGDRDGTGTRPS 266  
DB 123 VEEETHGFRWVEERDFTGIDGTENP-----QGDKRPEVAVIADG-----EEDAG---- 167  
QY 267 WALDGSFMAHRHFO-----QKYPEFNATLANAIPANSAGNLQOEGAEFLGARMFGRW 320  
DB 168 ---GSYVLVQRYEHLNKNQRIPE-----NEQE-----KILGT 198  
QY 321 KSGAPIDLAPTADDPALGADPQNNNFYSDTLTDTETCPFGAHVRKTNPRDLGSPVDT 380  
DB 199 K-----LDSQELPSD-QR-----PDTSHSVRDLKENGKNG----- 227  
QY 381 PHAMRSSIPYGPETSDAELASGVTAQDRGLLFVEY-----QSIIGNGFRFQOINWAN 432  
DB 228 LKILRQSLPYG-----LASG-----KHGLYFIAYCARLHNIEQQLLS---MFGDIDGKH 273  
QY 433 NANPFFSKPIPTPGIEPIIGQTTPTVGGDLPLNQNETTPTVPLVPKGEYFFLPSPISAL 492  
DB 274 DQLURFSKPV-----GSYYFAPSLTAL 296

RESULT 12  
Q8CGV6 PRELIMINARY; PRT; 1093 AA.  
AC Q8CGV6;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein ykbc.  
GN YKBC OR L10986.  
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
OC Streptococcaceae; Lactococcus.  
OX NCBI\_TaxID=1360;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IL1403;  
RX MEDLINE=21235186; PubMed=11337471;  
RA Bolotin A., Wincker P., Nausier S., Jaillon O., Malarne K.,  
RA Weissenbach J., Ehrlich S.D., Sorokin A.;  
RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
RT lactis ssp. lactis IL1403";  
RL Genome Res. 11:731-753(2001).  
DR EMBL; AE006333; AAK05084.1; --  
DR InterPro; IPR005044; RNA\_rec\_mot.  
DR PROSITE; PS00030; RRM\_RNF\_1; UNKNOWN 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 1093 AA; 116310 MW; 4615B4393PB50317 CRC64;

Query Match 4.7%; Score 121.5; DB 16; Length 1093;  
Best Local Similarity 21.3%; Pred. No. 2.7;

Matches 123; Conservative 64; Mismatches 228; Indels 163; Gaps 27;

```

OY 17 SSHVMAALGARQRTTP-LITNPGQAPLPTLTQ-HTTSGANDTILPLANNIOGDIIVG 74
DB 62 STYVNSLAASSTTEFAVVSNNMENTAUVSSSTNLTITSSSTVDNPSPTOSNOMIALS 121
OY 75 MKKOKERFVFVNDATSEFKALKTVPRTSAILISDPQOPLAFVNLGFSN----- 129
DB 122 LTKARESSINBPATINETTATADVTOTATQAOTSDPTTS-LSSSEQKPTNPSOS 180
OY 130 -----TGLOALGTTD--LGDAPFPDQGFADANLGDLSQWAPRTGTTTGVLIGSD 182
DB 181 KSEITNTIQTAVNNATGSAVF-----DGVNI-----TLGG----- 213
OY 183 ODDPFDQFDDISSTFGSSITVOVALSGSARFPDQAGHEHFGFLDGISQPSVTGMETTVF 242
DB 214 -KDITDNDNLDSGSHMSEOTOVAIKGTAT--GQLNQENFGVQDDPIIPATT--YTAN 267
OY 243 PQQA--VPPGIIITGRDQ-----TGTSPW-ALDGSMAFRHQKVPFNATYTLAN 293
DB 268 DDDSGRIYVGSLSGLDLMITYVASDDKDSQWQANGA-----EGIPQGLTFTTGEQ 319
OY 294 AIPANSAGN--LTOQGAFLGARMFGKSGAPIDLAPTAD-----DPALGAD----- 340
DB 320 NI--AASGGSIVCLANGNAL--SLIYQIVKHDTTEVPVVASPTTDDINAQGVQNTILAN 377
OY 341 -----PQRNNNFVSDTLTDET-----RCPFGAHR----- 366
DB 378 LVITLIPQTNLKQDDDTTYDASPNYPGLDGVASLPYGYLGAGFSEFYNNYAPAPER 437
OY 367 -----KTNPQDGLGAPDTPFAMRSSI-----PYGPETSDAELASGYTAQDRGL 410
DB 438 DDSYFPAQGVRYDLFGSLAQAH--MNTQIRQNFVNYVYDFGHKIQETDHLIGTQGDVYL 496
OY 411 LVEVQSIIIGNGF-----RFOQINMANNANFP-----FSKPITP 444
DB 497 ---PIPTIKGFSEFVNLTEMDASKNNPYNLILYNNLPTLYGNHNNNITYOGTAYTPSGFTI 553
OY 445 GIEPIIGQTPPTVGLDPLNQNETFVPLFVPIKGE 482
DB 554 GYQNI---GNPEASITYPVNNKASSYTLPLIMAGKK 588

RESULT 13
O9RK02 PRELIMINARY; PRT; 420 AA.
ID O9RK02
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative membrane protein.
GN SC02276 OR SCC75A.22.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Jarke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RL "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
DR EMBL; AL133220; CAB61722.1; -.
SQ SEQUENCE 420 AA; 44522 MW; 56344615594E01BC CRC64;
```

Query Match 4.6%; Score 119.5; DB 16; Length 420;  
 Best Local Similarity 21.9%; Pred. No. 0.89;  
 Matches 81; Conservative 35; Mismatches 121; Indels 133; Gaps 17;

```

OY 89 DATSEFKALKTVPRTSAILISDPQOPLAFVNLGFSNLTQALGTTDLGDA----- 144
DB 125 DDTGEMALGK--PSRLT-----LRTGFCPSLFTFGLADLRPEALDL 165
OY 145 -QPPQGFADANLGDLS--QWAPRTGTTIHGV-----FLGSGQDDFLQFDDISSTFG 199
DB 166 PKFP-GDNLDRARSGDLCVQACADDPQAVAHIRMLARTG-----FG 207
OY 200 SSTVOVAL-----SGSARFPDQAGHEHFGFLDGISQPSVTGMETTVFPQGA VPPGIIIT 255
DB 208 KYVWMSQLGFGKTSSTTPDEQTPRNLLGFGDGRN-----IA 245
OY 256 GRDQGTGRPSMAD-----GSPMARFRHQKVPFNATYTLANAI PANSAGNLTOQ 306
DB 246 GTEKORLDFVVAEKDQTPMWTGSSYLVARIRIMHETWD-----RASLQEQ 293
OY 307 EGAEFLGARMFGRWK--SGAPIDLAPTADDPALGA--DQRNNNFVSDTLTDET RCPFGAH 364
DB 294 ED-----VFGGRDKGEGAPVGVKAKERDEPFLVAMKPD-----AH 326
OY 365 VKRTPRODGLGAPDTPFAMRSSI-PYGPETSDAELASGYTAQDRGLL FVEVQSIIIGNGFR 424
DB 327 VRLAHPDSNGATL-----LRRGYSFTDGT-----DGLGRLDAGLFLAYORDIRTFV 375
OY 425 FQOINMANNA 434
DB 376 PVQRNLATDA 385
```

```

RESULT 14
O912M3 PRELIMINARY; PRT; 2468 AA.
ID O912M3
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Hypothetical protein PA1874.
GN PA1874.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINB=20437337; PubMed=10994043;
RA Hickey C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
RA Brody L.L., Goltser L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kae A., Lartig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004613; AAC05263.1; -.
DR InterPro; IPR001343; HemIysn_Ca_bind.
DR InterPro; IPR003880; Pantne_attach.
DR PRINTS; PR00313; CABINDGRPT.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 2468 AA; 258414 MW; 13596FBA2CB899 CRC64;
```

Query Match 4.6%; Score 118.5; DB 16; Length 2468;  
 Best Local Similarity 21.1%; Pred. No. 16;  
 Matches 115; Conservative 57; Mismatches 218; Indels 155; Gaps 28;

21 NAKKL-GAQTOTTTLTLPFGQAPLPTLTQHTTSGANDTILPLANNIOGDIIVGKKKK 79

```
Db 1088 NGSSLSGTAEPGTVILTDGNGN-----PIAEVTDAGSGNWTPTPPIANGTVVNVVAQ- 1142
QY 80 ERFVFFQVNDATSEKLTALKTVVQRIITSAAILISDPQOPLAFVN--LGFSNTGLQALGI 137
Db 1143 -----DASG-----NSPPATVTVDSAPPAPVINSNGVVISGTAEGA 1182
QY 138 TDLGDA-OFGDGPQADANLGDLSQWVAPFT-GTTI-HGVFLIGSDQDDELDFDFTDDI 194
Db 1183 TVTLTDAGNPIGQVT-----ADSGNM--SFTPGTPLANGTVIVAT-----ATOPT 1227
QY 195 SSTFGSSITQVQALSGSARPEQAGHEHFGFLDGISQPSVTGMETTVFPGQAVVPPGIIL 254
Db 1228 GNTGQQAATTVDVAPPAPVID-----PS-----NGTTISGTAEGAKVIL 1268
QY 255 TGRDGDGTRSWALDGSMAFRPQKQVPEFNAYTLANAI PANSAGNLTQOEGAEFL-- 312
Db 1269 T--DGNGNPIGETTADGS--GNWSEFTCTPLANG--TVNVAQDPAGN--TGPQSGTTTDA 1322
QY 313 -----GARMFERKSGAPIDLAPTADDPALGADPQNNNFYSDTLTDTETCP 360
Db 1323 VAPNTPVNVNPSNGNLLNGTAPGSGTVTLTDGNGNPIGOTTADGSGNWSF-----TPGSQLP 1378
QY 361 FGHRKYNPRQDLGG-----PVDTFHMRSGIPIYGPEP--SDAELASGVTADQRLGFVE 414
Db 1379 NGTVVNVMT--ASDAAGNTSLPAT--TVDSLSPLSIPOVDPSPNGSVISGTADAGNTIIITD 1434
QY 415 YQSIIGNGFRFOOI-----NWNANNAPPPS-----KP 441
Db 1435 -----GNGNPIGQVTADGSGNKSFTPGIPLPDGTVNVNVARSPSNVDSAPAVITVDGVAP 1489
QY 442 ITPGIEPI-----IGQTPRTVGGDLPLNQNETHFTVPLFVLPKGGEYFFLPSISA 491
Db 1490 AAPVIDPSNGTEISGTAEGAGATVILTDGNGNPIGQ-----ATADGSGNWTFTPTPPL 1541
QY 492 LTATI 496
Db 1542 ANGTV 1546
```

## RESULT 15

```
Q9L800
ID Q9L800 PRELIMINARY; PRT; 2747 AA.
AC Q9L800;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE RTX protein.
GN ASX.
OS Aeromonas salmonicida.
OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
OC Aeromonas.
OX NCBI_TaxID=645;
RN [1]_TaxID=645;
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33656;
RA Braun M., Frey J., Kuhnert P.;
RT "280 kDa RTX protein of Aeromonas.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF218037; AAF27914.1;
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR001343; Hemlyan_Ca_bind.
DR InterPro; IPR003410; Hyalin.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00353; hemolysinCbind; 9.
DR PRINTS; PR00313; CABNDNGRPT.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00130; HEMOLYSIN_CALCIIUM; 4.
DR PROSITE; PS50234; VVFA; 1.
SQ SEQUENCE 2747 AA; 280201 MW; 208FE380E44A5F37 CRC64;
```

Query Match.

4.5%; Score 118; DB 2; Length 2747;

```
Best Local Similarity 19.7%; Pred. No. 20;
Matches 104; Conservative 57; Mismatches 204; Indels 162; Gaps 22;
QY 32 TTPLLTNFPQAPLPTLTQHTTESGANDTILPLNNIQGSDILVGMKKQKRFVFFQVNDAT 91
Db 288 TTPTTGTGTTDAAGSSTVTLVVTDANGNQQTLSAT-----VQPNG 326
QY 92 SEKTAKYTVVQRIITSAAILISDPQOPLAFVNLSGNTGLQALGITDLDGAPPDQGF 151
Db 327 TVSTDVTSPLPDGVDYVTVASVTDPAAGTGTATDGGSDVDTAPTALT--LDANITDDDV 383
QY 152 ADAANLGDLSQWVAPFTGT-----TIHG-----VFLIGSDQD 184
Db 384 INAAEAGQOI-----PVSGTVSGEFKAGDTVTLTVNGKPTGPDVADGRFFILVAGSD-- 436
QY 185 DFLDQFTDDISSTFGSSITQVQALSGSARPPDOAGHEHFGFLDGISQPSVTGMETTVPPG 244
Db 437 -----LAADTORTIDASVTSTDAAGNSATATDSEGY-----GVDTTAPVD 476
QY 245 QAV-----VPPGIILTG--RDGDTGT-----RPSWA---LDGSFM 274
Db 477 LAITLDANITDDDDVINAAEAGQOI PVSGTVGFEKAGDTVTLTVNGKPTGPDVADGRFT 536
QY 275 APRHFQKQVPEFNAYTLANAI PANSAGNLTQOEGAEFLGARMFGRWKSGAPIDLAPT--- 331
Db 537 ILVAGSDLAADTDRITIDASVTSTDAAGNSATATDSEGYGV-----DTTAPVDLAITLDA 590
QY 332 --ADDPALGADPQNNNFYSDTLTDTETCPFGAHRKYNPRQDLGGPVD-----TPHAM 384
Db 591 NITDDDVINA-AEAGQOI PVSGTVSGEFKA--GDTVTLTVNGKPTGPDVADGRFTILVA 647
QY 385 RSSIPYGPETSDAELASGVTADQRLGLLFVEYQSIQNGFRFQOINWANNANPPPS----- 439
Db 648 GSDL---AADSRTIDASVTSTDA--GN-----STTKSDTENYTVDTVAPD 689
QY 440 -----KPITPGIEPIIGQT-----TPRTVGGI--DPLNQNETHFTVPL 474
Db 690 LGIDLDPLAVGSDNVNVAEADGKTPVTLTSGTVNGDANVGDVTVTLTL 736
```

Search completed: November 27, 2002, 13:47:49

Job time : 41 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using bw model

Run on: November 27, 2002, 13:45:31 (Search time 16 Seconds  
(Without alignments)  
915.786 Million cell updates/sec

Title: US-09-926-084-7

Perfect score: 2601  
Sequence: 1 MDLSLFVSVAVLVGSSSHV.....KGGYFLFSLSALTATTTAA 498

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 2942292 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents-AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	112.5	4.3	1837	3	US-08-926-361B-5 Sequence 5, Appli
2	109.5	4.2	4928	4	US-09-036-987A-5 Sequence 5, Appli
3	109.5	4.2	4928	4	US-09-370-700-5 Sequence 5, Appli
4	102	3.9	2089	1	US-08-418-893D-23 Sequence 23, Appli
5	102	3.9	2089	1	US-08-418-893D-24 Sequence 23, Appli
6	101.5	3.9	2089	1	US-08-418-893D-24 Sequence 23, Appli
7	100	3.8	1003	2	US-07-743-357-10 Sequence 10, Appli
8	98.5	3.8	998	2	US-08-415-788-7 Sequence 7, Appli
9	98.5	3.8	1004	2	US-07-743-357-7 Sequence 7, Appli
10	98.5	3.8	1029	2	US-08-415-788-3 Sequence 7, Appli
11	97.5	3.7	913	2	US-07-743-357-22 Sequence 22, Appli
12	97.5	3.7	1005	2	US-07-743-357-1 Sequence 1, Appli
13	97	3.7	1003	2	US-07-743-357-9 Sequence 3, Appli
14	96	3.7	667	4	US-08-485-355B-50 Sequence 50, Appli
15	96	3.7	647	4	US-09-194-613-3 Sequence 3, Appli
16	96	3.7	675	4	US-08-485-355B-52 Sequence 52, Appli
17	95.5	3.7	1016	2	US-07-743-357-5 Sequence 5, Appli
18	95	3.7	490	4	US-09-252-149B-26 Sequence 26, Appli
19	95	3.7	699	2	US-08-694-965-16 Sequence 16, Appli
20	95	3.7	699	2	US-08-124-491-16 Sequence 16, Appli
21	95	3.7	1003	4	US-09-309-572-17 Sequence 17, Appli
22	94	3.6	1003	4	US-07-743-357-8 Sequence 8, Appli
23	93.5	3.6	414	4	US-09-715-834-2 Sequence 2, Appli
24	93.5	3.6	1381	4	US-08-540-245A-16 Sequence 16, Appli
25	92.5	3.6	501	2	US-08-913-477-23 Sequence 23, Appli
26	92.5	3.6	1016	2	US-07-743-357-3 Sequence 3, Appli
27	92.5	3.6	1042	3	US-08-928-361B-11 Sequence 11, Appli

28	91	3.5	316	1	US-07-828-980A-2	Sequence 2, Appli
29	91	3.5	1183	4	US-09-134-001C-3530	Sequence 3530, Ap
30	90	3.5	299	6	5514590-4	Patent No. 5514590
31	90	3.5	661	2	US-08-525-742-2	Sequence 2, Appli
32	89.5	3.4	651	2	US-08-687-702-37	Sequence 37, Appli
33	89.5	3.4	670	4	US-09-328-501-15	Sequence 15, Appli
34	89.5	3.4	1721	3	US-08-700-651-5	Sequence 5, Appli
35	89.5	3.4	1721	3	US-08-928-361B-6	Sequence 6, Appli
36	88.5	3.4	913	2	US-07-743-357-6	Sequence 6, Appli
37	88	3.4	513	4	US-09-071-035-68	Sequence 68, Appli
38	88	3.4	522	4	US-09-071-035-68	Sequence 68, Appli
39	88	3.4	746	5	PCT-US95-10509-2	Sequence 66, Appli
40	88	3.4	1140	4	US-09-220-081-2	Sequence 2, Appli
41	88	3.4	1140	4	US-09-677-575-2	Sequence 2, Appli
42	87.5	3.4	5215	4	US-09-105-537-2	Sequence 2, Appli
43	87	3.3	529	1	US-08-484-815-12	Sequence 12, Appli
44	87	3.3	529	3	US-08-888-949-12	Sequence 12, Appli
45	87	3.3	529	4	US-08-888-950-12	Sequence 12, Appli

## ALIGNMENTS

RESULT 1  
US-08-928-361B-5  
Sequence 5, Application US/08928361B  
Patent No. 6071318  
GENERAL INFORMATION:  
APPLICANT: Petersen, Carolyn  
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,  
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS  
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM  
TITLE OF INVENTION: SPECIES INFECTIONS  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSER: PETERS, VERNY, JONES & BIKSA  
STREET: 385 Sherman Avenue, Suite 6  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94106-1840  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/928,361B  
FILING DATE: 12-SEP-1997  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 60/026,062  
FILING DATE: 13-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: VERNY, Hana  
REGISTRATION NUMBER: 30,518  
REFERENCE/DOCKET NUMBER: 480,76-1(HV)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-324-1677  
TELEFAX: 650-324-1678  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1837 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-928-361B-5

Query Match 4.3%, Score 112.5, DB 3, Length 1837;  
Best Local Similarity 18.5%, Pred. No. 0.12;  
Matches 108; Conservative 64; Mismatches 168; Indels 245; Gaps 30;

```
QY 29 QTRTLLTNFGQAPL-----PTLQHTTSGANDTIL 62
| | | | |
DB 736 OKPTTTTTTKVGPPIATTTTTLKPIVITTTTTKATTTTTTTPTTTTTKRDEMTTIT 795
| | | | |
QY 63 PLNNLOGDI---LVCKMKKERF--VFQVNDATSPKALTAKTVVPORITSAAILLSDPSQ 117
| | | | |
DB 796 PLPDI-GDIEITPIEKMLDYRMIDYNSG-----LLDSDND 834
| | | | |
QY 118 QPLAFVNLGFSNTGLQALGITDLDGDAQPDGQFADANL----- 157
| | | | |
DB 835 EPI-----PGSQA-----GQIADTSNLFVQTHKSTGLPIDPMVGLP 871
| | | | |
QY 158 -GDDLQWVAPPTGTTIHGV-----FLIGSDQ----- 183
| | | | |
DB 872 FDPKGNLVHPYNTGMSGLSVSLAAKNLTVOTDETYGLPIDTLTGYPDPSVLIIPNP 931
| | | | |
QY 184 --DQDLQDFTDI--SSTFGSITVQA-----LGSARPFDQAGHEHFGFLDGISQPSVTG 236
| | | | |
DB 932 ETGELFDPISDEIMNGITAGIVSGISASELLSOKGAPIDPATNMVVGFEFGLLNP-TG 990
| | | | |
QY 237 WETTVPFGQ-----AVVPGIILTGRDGDGTGRPSWALDGSFMAFRHFQ 280
| | | | |
DB 991 ---VMIPGSLGPSEQTPSPRIEDGGIIPPEVAANADKFLSIPSPVE-----SIPEKD 1043
| | | | |
QY 281 QKVPEFN--AYTLANA-----IPANSAGNL-----TQEGAGFLGARMFGRWK 321
| | | | |
DB 1044 QKIDSISELMVDIESGRLIGQVSKRPPIPGSIAGDLNPIMKPTQTDSV----- 1091
| | | | |
QY 322 SCAPIDLAPTAD--DPALG--ADPQNNNFQYSDTLTDTRCPCGAH-----VRXTN 369
| | | | |
DB 1092 TKRPID--PTTGLPFPPTGHLINPTNNNTWDSFA-----GAYKAVSNGIKTDN 1140
| | | | |
QY 370 PRODLGGPVDTFHAMR-----SSIPYGPETSDA-----ELASGVTAQDRGLLFVEY 415
| | | | |
DB 1141 V---YGLPVEITGLPKDPVSDIIPFNSTTGTGELVDPSTGKPINNYTAGIVSGRKLPIED 1197
| | | | |
QY 416 OSIIGNFRFOGINWANNANPFPSKPIPTPGIEPIIGOTTPRTVGG 460
| | | | |
DB 1198 E-----NGNLFD-----PSTKLPIDGNQLNVPETNSTVSG 1228
| | | | |

RESULT 2
US-09-036-987A-5
; Sequence 5, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merio, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
```

```
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4928 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-036-987A-5

Query Match 4.2%; Score 109.5; DB 4; Length 4928;
Best Local Similarity 23.5%; Pred. No. 1.2;
Matches 100; Conservative 40; Mismatches 169; Indels 117; Gaps 24;

QY 30 TTTTLLTNFGQAPLPTLTQHTTSGANDTILPLNNIOGDIIVGMKKQKRFVFFQVND 89
| | | | |
DB 3190 SRRRLDDELPEAQALADAENTTD--AADSAVPLPR-----LAGMAAAERR----- 3234
| | | | |
QY 90 ATSFKTALKTVVPORITSAAILI-----SDPSQOPLAFVNLGFSNTGLQALGITDLDGA- 144
| | | | |
DB 3235 -----RAMLDLVLAESIVLGHNGSDPVGPDRAFOELGPDFS--LMAVELRNRLGEAT 3284
| | | | |
QY 145 --QFPDQGFADAAAGLDLQWVAPPTGTTIHGVFLIGSDQDDFLDQFTDDISSTFGSSI 202
| | | | |
DB 3285 GLSLPATLIFDYPSPSALAEQLGELVGAPATTVVAGADPVD-----DPV----- 3330
| | | | |
QY 203 TVQALSSASRPFOAGHEHFGFL-----DGISQ--PSVTGWE--TTVF-----PQAVVP 249
| | | | |
DB 3331 --VVVAMGCRYPGVCSPEBLWQLVSGRADAVSTFPVDRGMDCNLTFDPDPRAGSTYVR 3388
| | | | |
QY 250 PGIIITGRDG--DTG-----TRPSWALDGSFMAFRHFQOKV-----PEFNAYTLANAIPANSA 300
| | | | |
DB 3389 EGAFITGADREFDAGFFGISPREARMDP-----QQRLLLEVAWEVFERAGIAPLSLR 3440
| | | | |
QY 301 GMLTQEGAEFLGARMFGRWKSGAPIDLAPTADDPALGADPQNNNFQYSDTLTDTRCPC 360
| | | | |
DB 3441 GSRT-----GVFAGTN--GQ-DHGAKVAAAPEAAGHLLTGNA-----ASVLAGRLSYT 3485
| | | | |
QY 361 FGAHVKTNPQRDLGGP---VDT-----PHAMRSSIPYGPETSDAELASGVTAQDRG 409
| | | | |
DB 3486 FG-----LEGFAVAVDTACSSSLVALHLACQSLRSG--ECDNALAGGVTVNSTP 3532
| | | | |
QY 410 LLFVEY 415
| | | | |
DB 3533 LAFLEF 3538

RESULT 3
US-09-370-700-5
; Sequence 5, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 Div1
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
```



LENGTH: 4928  
 TYPE: PRT  
 ORGANISM: Saccharopolyspora spinosa  
 us-09-370-700-5

Query Match 4.2%; Score 109.5; DB 4; Length 4928;  
 Best Local Similarity 23.5%; Pred. No. 1.2;  
 Matches 100; Conservative 40; Mismatches 169; Indels 117; Gaps 24;

30 TTTPLTNFPGQALPPLTQHTTESGANDTILPLNNIQQDILVGMKKKQKRRVFVQVND 89  
 3190 SRRRPLDLPLPAQQLADAENTTD--AADSAPLPR-----LAGMAAARR----- 3234  
 90 ATSFKTALKTVPQRITSAAILI-----SDPSQOPLAFVNLGFSNTGLQALGITDLDGA- 144  
 3235 -----RAMLDVLAESAYLGHNGSDPVGPDRAPFOELGFDG--LMAVELRNRLGEAT 3284  
 145 --QFPGGFADANIGDDLSQWVAFPTGTTIHGVFLIGSDODDFLDQFDIDISTFGSSI 202  
 3285 GLSLPRTLIFDPSFALAEQLVGLVGAQPAITTVAGADVD-----DPV----- 3330  
 203 TOYQALSGARPFQDAGHEHFGFL-----DGIQ-PSVTGME-TTVE-----PGQAVVP 249  
 3331 --VVVAMGCRVYGDVCSPELWQLVASGRDAVSTFPVDRGDCDCTLFPDPDRAGSTYR 3388  
 250 PGILGRCG-DYG-----TRPSMALDGSFMAFRHFOQV---PEFAYTLANAIPANSA 300  
 3389 EGAFELGADRFAGFFGIGSPREARANDP-----QORLLLEVAMEVFERAGIAPLSLR 3440  
 301 GNLTOGEAFLGARMFGKSGAPIDLAFTADDPALGADPORNNDVSDTLTDETRCP 360  
 3441 GSRT-----GVFAGTN--CG-DHGAKVAAAPRAGHLTLGNA-----ASVLGRLSYT 3485  
 361 FGHNVRKTPRODLGSP--VPT-----FHMRSSIPYGPSTDAELASVTADQDG 409  
 3486 FG-----LGGPAVAVDTACSSSLVALHLACGSLRSG--ECDMALAGAVTWMSTP 3532  
 410 LLEVEY 415  
 3533 LAFLER 3538

RESULT 4  
 US-08-418-893D-23  
 Sequence 23, Application US/08418893D  
 Patent No. 5559220  
 GENERAL INFORMATION:  
 APPLICANT: ROESSLER, PAUL G  
 APPLICANT: OHROGGE, JOHN B  
 TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A  
 TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTHELLA CRYPTICA  
 NUMBER OF SEQUENCES: 25  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY  
 STREET: 1617 Cole Blvd.  
 CITY: Golden  
 STATE: CO  
 COUNTRY: USA  
 ZIP: 80401-3393  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/418, 893D  
 FILING DATE: April 7, 1995  
 CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/104,938  
 FILING DATE: September 14, 1993  
 CLASSIFICATION: 800  
 ATTORNEY/AGENT INFORMATION:

NAME: O'CONNOR, EDNA  
 REGISTRATION NUMBER: 29,252  
 REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 303-231-1000  
 TELEFAX: 303-231-1098  
 TELEX:

INFORMATION FOR SEQ ID NO: 23:  
 SEQUENCE CHARACTERISTICS:

LENGTH: 2089 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 US-08-418-893D-23

Query Match 3.9%; Score 102; DB 1; Length 2089;  
 Best Local Similarity 21.0%; Pred. No. 1.6;  
 Matches 88; Conservative 53; Mismatches 168; Indels 110; Gaps 21;

9 SVALVSGSSHNNAKLGAQRTTPTPLTNFPGQALPPLTQHTTESGANDTIL-----PL 64  
 1204 NVQIYVGS---KTRARRAAQVFLRAISHPTGL-----TTFSGARRALLQGLDEL 1251  
 65 NNIQDILVGMKKKQKRRVFP--QVNDATSFKTR-----LXTYVFORI----- 105  
 1252 ERAQANSKVSVQSSRIYLHSLPEQSDATPEEIAKEFEGYIDLXKRLAQRLKLRVDEI 1311  
 106 -TSAAILISD-----PSQOPLAFVNLGFSNTGLQALGITDLDGACQPDQFADAAATLGD 160  
 1312 ETQVRYTVDEDEGSPRVVPRVLVASMOWGELKTSAYIDR-----PD----- 1353  
 161 LSGWVAFPTGTTIHGVFLIGSDODDFLDQFTDDISTFGSITQVALSGSARPFQAG- 219  
 1354 -----PVTGTVRERC-VIGGIDVCELESYSTSTIOTK-RSIRARVSTYAVDYGL 1405  
 220 -----HEHFGFLDGISOPSVTGWETVFPQAVV--PGIILTGRDGTGRP-----SW 267  
 1406 LEVSLGEMDKLSSLSGPDPTPIPSNVFEAQELBEPDDELVTGK-REIGTNKGMVAM 1464  
 268 ALDGSFMAFRHFOQVPER-----NAYTLANAIPANSAGNLTOGEAFLGARMFG- 319  
 1465 VV-----TMKTPPEYRPGRGVVYVNDVTVQS--GSFGVEDEDFKASKYARENKL 1513  
 320 -----WKSAPI-----DLAPTADDPALG--ADPORNNDVSDTLTDETRCPGAVRK 367  
 1514 PRVYINCNGAIGLVDDLKRFQIKFIDESPCKGEFVLYDDATYKSLPESGVNVRK 1572

RESULT 5  
 US-08-418-893D-24  
 Sequence 24, Application US/08418893D  
 Patent No. 5559220  
 GENERAL INFORMATION:  
 APPLICANT: ROESSLER, PAUL G  
 APPLICANT: OHROGGE, JOHN B  
 TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A  
 TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTHELLA CRYPTICA  
 NUMBER OF SEQUENCES: 25  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY  
 STREET: 1617 Cole Blvd.  
 CITY: Golden  
 STATE: CO  
 COUNTRY: USA  
 ZIP: 80401-3393  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/418,893D  
 FILING DATE: April 7, 1995  
 CLASSIFICATION: 800  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/104,938  
 FILING DATE: September 14, 1993  
 CLASSIFICATION: 800  
 ATTORNEY/AGENT INFORMATION:  
 NAME: O'CONNOR, EDNA  
 REGISTRATION NUMBER: 29,252  
 REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 303-231-1000  
 TELEFAX: 303-231-1098  
 TELEX:

INFORMATION FOR SEQ ID NO: 24:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 2089 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 FRAGMENT TYPE: N-terminal  
 US-08-418-893D-24

Query Match 3.9%; Score 102; DB 1; Length 2089;  
 Best Local Similarity 21.0%; Pred. No. 1.6; 168; Indels 110; Gaps 21;  
 Matches 88; Conservative 53; Mismatches 110; Gaps 21;

QY 9 SVAVLVGSSSHVNAKLGARQTRTPLLTFPGQAPLPTLTOHTTESGANDTIL-----PL 64  
 DB 1204 NVOIVGSE--KTARRNAAQVFLRAISHTPGL-----TTFSGARRALLOGLDEL 1251  
 QY 65 NNIQDILVGMKKKERVFF--QVNDATSEKTA-----LKVYVQRI----- 105  
 DB 1252 ERAQANSKVSVSSRIYHLSLPSQSDATPEETAKEFEGVIDKLKSLRAQRLTKLRVDEI 1311  
 QY 106 -TSAAILSD-----PSQOPLAFVNLGFSNTGLQALGITDLDGDAQFPDQGFADANLGD 160  
 DB 1312 ETKVRVTQDEGSPRVVFLVASSWQGEWLKTSAYIDR-----PD----- 1353  
 QY 161 LSQWVAPFTGTHGVFLIGSDDDDFDQFTDDISSTFGSSITQVQALSSARPFDAQ- 219  
 DB 1354 -----PVTGVTRRC-VIGGIDEVELESYSTSTIQTK-RSIARRVGSYAYDYGL 1405  
 QY 220 -----HEHFGFLDGISQPSVTGNETTVFPQAVV--PPGILTGDRDGTGRP-----SW 267  
 DB 1406 LEVSLLGWDKYLSSLGPDPTTIPSNVFEAQELLEGPDGELVTGK-REIGNTKVGMVAM 1464  
 QY 268 ALDGSFMAFRHFOOKVPEF-----NAYTLANAIPANSAGNLTOEGAEFLGARMFGR----- 319  
 DB 1465 VV-----TMKTPYPRGROVVVINDVTVQS--GSPGVEEDVFFKASKYARENKL 1513  
 QY 320 -----WKSQAPI-----DLAFTADDPALG-ADPQRNNNFYDSTLTDETRCPFGAHVRK 367  
 DB 1514 PRVVIACNSGARIGLVDLKKPKQIKTIDEASPSKGFVLYLDATYKSLPEGSVNVVRK 1572

RESULT 6

Patent No. 5268290-2  
 APPLICANT: Hasegawa, Mamoru; Sakurada, Kazuhiro  
 TITLE OF INVENTION: PROCESS FOR PRODUCING NEURAMINIDASE  
 NUMBER OF SEQUENCES: 2  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/384,709  
 FILING DATE: 25-JUL-1989  
 SEQ ID NO: 2:  
 LENGTH: 626  
 5268290-2

Query Match 3.9%; Score 101.5; DB 6; Length 626;  
 Best Local Similarity 20.2%; Pred. No. 0.27;  
 Matches 102; Conservative 45; Mismatches 145; Indels 213; Gaps 28;  
 QY 116 SOOPLAFVN--LGSNTGLQALGITDLDGDAQFPDQGFADANLGDLSQWVAPFTGTTI 173  
 DB 15 TEQDLA-VNREGFPNRIIPALVTTPD-----GDLIASYDGRPTGIDA 56  
 QY 174 HGVFLIGSDDDDFDQFTDDISSTFGSSITQVQALSSARPFDAQHEHFGFLDGISQPS 233  
 DB 57 PG-----PNSILQRRSTGGRTWGEQ--QVVSAGOTTAP-----INGFSQPS 96  
 QY 234 -VTGHET--TVF-----PGQAVVPPGI-----ILTGRDG----- 259  
 DB 97 YLVDRGTGTIFNFHVYSORQGFAGSRPGTDPADPNVLHANVATSTGGTGLTWSHRTITADI 156  
 QY 260 --DTCTRPSSWALDGSFMAFRHFOOKVPEFNAYTLANAIPANSAGNLTOEGAEFLGARMF 317  
 DB 157 TPDPCWRSRFAASGEGQLRVGPHAGRLIQQYTIINAACAPQAVSVSDD-----H 207  
 QY 318 GR-WKSGAPIDLAFADDPALGADPQNNNFYSD-----TLTDETRCPFGAHVRKTNPR 371  
 DB 208 GRTWAG-----EAVGMDENKTVELSDGRVLLNSRDSARSGY-----RKVAVS 252  
 QY 372 QDLGCFVDTFHMRSSIPYGPETSDAEL-----ASGVTAAQDRGLLF----- 412  
 DB 253 TD-GG-----YGPVTIDRDLDPDTNNASIIIRAPDAPAGSARAKVLLFSNAA 300  
 QY 413 -----VEQSI-----GNGRPFQOIN 429  
 DB 301 SOTSRQGTIRMSDDQGTWPVKVFGQSMYSYTLTALPDGTYGLLYEPGTGTRYANFN 360  
 QY 430 --WANNANFPFSK-----ITPGIEPII-----GQTPR-----TVGGLDPL 464  
 DB 361 LAWLGGICAPPTIPDVALEPGQVTVVAVTNQSGTAVPKPSLQLDASPDQVQGSVEPL 420  
 QY 465 NONE-----TFTVPLFVIPKGEY 483  
 DB 421 MFGROAKGQVTTITVPAGTTP--GRY 443

RESULT 7

US-07-743-357-10  
 Sequence 10, Application US/07743357  
 Patent No. 5858646  
 GENERAL INFORMATION:  
 APPLICANT: Kang, Yong C.  
 TITLE OF INVENTION: Polypeptide having immunological  
 TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
 NUMBER OF SEQUENCES: 22  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: KIRBY EADES GALE BAKER  
 STREET: Box 3432, Station D  
 CITY: Ottawa  
 STATE: Ontario  
 COUNTRY: Canada  
 ZIP: K1M 1H8  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patencin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/743,357  
 FILING DATE: 21-AUG-1991  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/CA90/00062  
 FILING DATE: 23-FEB-1990  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Gale, Edwin J.  
 REGISTRATION NUMBER: 28,584

REFERENCE/DOCKET NUMBER: 30924-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (613) 237-6900  
TELEFAX: (613) 237-0045  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1003 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
ORGANISM: Human immunodeficiency virus type 1  
STRAIN: ELI  
US-07-743-357-10

Query Match 3.8%; Score 100; DB 2; Length 1003;  
Best Local Similarity 22.1%; Pred. No. 0.8;  
Matches 76; Conservative 41; Mismatches 117; Indels 110; Gaps 18;

85 FQVNDATSKTL--KTYVPRITSAALISDPSCQ--PLAFVNLGFSNTGLQALGITDD 140  
Db 594 FVYDGAANEITLKGAGVYTDGRKXVPLDTNQLQALALQDSGLVNIIVTDS 653  
Qy 141 --LGDAP-PRGQPADAN-LGDDLQ-----WAPFTGTTIGCVFLIGSDQDFLD 188  
Db 654 QVALGIIQAQPKRSEBELNQIIEQLIKKYLAWPAHKG-----IGGNEQVDKL- 705  
Qy 189 QFTDDISFTGSSITVOALSGSARFPDQGEH-----FGFLDGISPSVT 235  
Db 706 -----VSGGIRKVLFLDG-----IDKQGEHEKXNNMRAASPNLPVVAEIVA 752  
Qy 236 GWETTVFPGQV-----VPPGIIITGRDGTGTRPSMALDGSFMAFHPQOKYPERNAIT 290  
Db 753 SCRCOLKEAMHGOVDSPGI-----WOLDCTHLEGVILVAHVASGYI 798  
Qy 291 LANAIPIANGAGNLTOEGAEFLGARMFGKSGAPIDLAPADPALGADPQRNNPDVS 350  
Db 799 EAEVIRAE-----TOGETHYFL-LKLAGRW-----PVKVVH-----DNGSNF----- 835  
Qy 351 DTLTDETRCPFGAHVKTNPRODLGGPVDTFHMRSSIPYGPET 394  
Db 836 -----TSAVVAKACWMAIGIKQEPG-----IPYNPOS 861

RESULT 8  
US-08-415-788-7  
Sequence 7, Application US/08415788  
Patent No. 5834351  
GENERAL INFORMATION:  
APPLICANT: NORMARK, STAFFAN  
APPLICANT: JONSSON, ANN-BETH  
TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES USEFUL  
NUMBER OF INVENTIONS: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/415,788  
FILING DATE:  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/829,465  
FILING DATE: 31-JAN-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: SCHWENNING, LYNN E.  
REGISTRATION NUMBER: 37,233  
REFERENCE/DOCKET NUMBER: 29500-20046.20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 998 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-415-788-7

Query Match 3.8%; Score 98.5; DB 2; Length 998;  
Best Local Similarity 20.5%; Pred. No. 1.1;  
Matches 120; Conservative 55; Mismatches 200; Indels 209; Gaps 30;

24 KLGARQRTPLTLTPPG-----QAPLPTLQHTTSG-----ANDTILPLNNI 67  
Db 279 KLGAPQVK-AGRCTKNPNNNNTKAPSPALPAMFGPGDGAEMYSASVSTYDSSS 337  
Qy 68 QGDILVMKKQK-----RVEFQVNDATSKTALKTYPQ-----RITSAILI-- 112  
Db 338 SRIFLQELKTQTEPGKPGYSLKSLNDG-----BIKRSQSPFNQRITIRLDGVHLIKL 392  
Qy 113 SDPSQQLAFVNLGFSNTGL-QALGITDDLDAQFPDQGFADANLGGDLSQVAFPTGT 171  
Db 393 NSKDEVAAPFVNLNGNNTKADTFEIVK-----ANVLDLDEKKVLLRW 438  
Qy 172 TIHGVFLIGSDQDFLQFTDDISFTGSSITVOALSGSARFPDQGEHFGFLDGISQ 231  
Db 439 TVRG-----PUNDKRFKSIHQKEKYSQRYRIRDNNNGNDLG--DIVNS 480  
Qy 232 P--SYTGWETTV-----PFG-----QAVVPPGII 253  
Db 481 PIYVAGSILATANDQWHIFKNGSDERSYNLKSITIEPMRKIQOSSTLAKELR 540  
Qy 254 LTRGRDGTGTRPSMALDGSFNAFR--HFQGXVEF-----NAVTLANAIPIANGAGN 302  
Db 541 AFAEKGVYQDR--YGVGGGFVLRQVELSGQKHVFMFGAMGPGRGAYAL-----D 588  
Qy 303 LTOEGAEFLGABMF-----GRWKSAPIDLAPADDPALG----- 338  
Db 589 LSKINGNYPAAPLFDVQDGNNGNRYKVELGTVGTPOIGIKRNGKYAFLASGYAK 648  
Qy 339 -ADPQRNN-----FDVSDTLTDETRCPFGAHVKTNPRODLGFPVDTFHMRSSIPYGP 393  
Db 649 KIDDSNKTALVYDILKDTL-----GTPIAKIEVDKGG-----LSS-----PT 688  
Qy 394 TSDAEIASGT--ADNRGLVVEQSIIGNGPFOQIN-----WANNANFPSPKPIITG 445  
Db 689 LVNDKLDGTDVDAIVAGDRG-----GNNYFDDLSNDSGSKWAKVIFEDKPIITSA 738  
Qy 446 IEBIIQTTPTV-----GSL--DPINONEFTVPLFVPIPKG 480  
Db 739 --PAVSRLABKRVYFGTSDLTEDDVANTGEQYIYGFDDDKG 780

RESULT 9  
US-07-743-357-7  
Sequence 7, Application US/07743357  
Patent No. 5858646  
GENERAL INFORMATION:  
APPLICANT: Kang, Yong C.  
TITLE OF INVENTION: Polypeptide having immunological  
NUMBER OF INVENTIONS: 22

```
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: KIRBY EADES GALE BAKER
/ STREET: Box 3432, Station D
/ CITY: Ottawa
/ STATE: Ontario
/ COUNTRY: Canada
/ ZIP: K1M 1H8
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/07/743,357
/ FILING DATE: 21-AUG-1991
/ CLASSIFICATION: 424
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: PCT/CA90/00062
/ FILING DATE: 23-FEB-1990
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Gale, Edwin J.
/ REGISTRATION NUMBER: 28,584
/ REFERENCE/DOCKET NUMBER: 30924-2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (613) 237-6900
/ TELEFAX: (613) 237-0045
/ INFORMATION FOR SEQ ID NO: 7:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1004 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: not relevant
/ TOPOLOGY: not relevant
/ MOLECULE TYPE: protein
/ HYPOTHETICAL: NO
/ FRAGMENT TYPE: internal
/ ORIGINAL SOURCE:
/ ORGANISM: Human immunodeficiency virus type 1
/ STRAIN: SF2
/ US-07-743-357-7

Query Match 3.8%; Score 98.5; DB 2; Length 1004;
Best Local Similarity 22.0%; Pred. No. 1.1;
Matches 78; Conservative 49; Mismatches 114; Indels 113; Gaps 20;

QY 78 QKERFV---FFQVNDATSFYAL--KTYVPQRITSAAILSDPSQ--PLAFVNLGFSNT 130
DB 585 EKEPIVGAETFYVDGAANRETKLGKAGYVDRGRQKVSIADTTNOKTELOAIHLALQDS 644
QY 131 GLQALGHTDD---LGDAQF-PGQOFAD-AAANLGDLLSQ-----WVAPFTGTTIHGVFL 178
DB 645 GLEYNIVTDSQYALGIIQAQPKRSELSVQIIEQLIKKEKVLAWVPAHKG-----I 697
QY 179 IGSDDDFLDOFTDDISSTGSSITQVALSGSARPFDOAGHEH-----PGF 225
DB 698 GGNBOVDKL-----VSAGIRKVLFLNG-----IDKAEHEKYSNWRAMASDFNL 743
QY 226 LDGISQPSVGTWETVTFPGQAV-----VPPGIILTRGDGTGTRPSWALDGSFMAFRHQ 280
DB 744 PPVVAKEIVASCDKQLKGEMHQVDCSPG1-----WQLDCTHLEKGIIL 789
QY 281 QKVPEFNAYTLANAIPANSAGNLTQOEGAEFLGARMFGRWKSAGPIDLADPTADDPALCAD 340
DB 790 VAVHVASGYIEAEVIPAE-----TGOBTAYFL-LKLAGRW-----PVKTIHT----- 830
QY 341 PQRNNDYSDTLTDETRCPFGARVRKTRNPQDLGGPVDTFHMRSSIPYGPET 394
DB 831 ---DNGSNFTST-TVKAACW-----AGIKQEBF-----IPYNPOS 862

RESULT 10
US-08-415-788-3
; Sequence 3, Application US/08415788
; Patent No. 5834591

/ GENERAL INFORMATION:
/ APPLICANT: NORMARK, STAFFAN
/ ADDRESSEE: JONSSON, ANN-BETH
/ TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES USEFUL
/ TITLE OF INVENTION: FOR THE DIAGNOSIS AND TREATMENT OF PATHOGENIC NEISSERIA
/ NUMBER OF SEQUENCES: 43
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: MORRISON & FOERSTER
/ STREET: 755 Page Mill Road
/ CITY: Palo Alto
/ STATE: California
/ COUNTRY: USA
/ ZIP: 94304-1018
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/415,788
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/829,465
/ FILING DATE: 31-JAN-1992
/ ATTORNEY/AGENT INFORMATION:
/ NAME: SCHWENNING, LYNN E.
/ REGISTRATION NUMBER: 37,233
/ REFERENCE/DOCKET NUMBER: 29500-20046.20
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 813-5600
/ TELEFAX: (415) 494-0792
/ TELEX: 706141
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1029 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-415-788-3

Query Match 3.8%; Score 98.5; DB 2; Length 1029;
Best Local Similarity 20.5%; Pred. No. 1.2;
Matches 120; Conservative 55; Mismatches 200; Indels 209; Gaps 30;

QY 24 KLGARQRTPTPLLTNPGF-----QAPLPTLTQHTTESG-----ANDTILPLNNI 67
DB 310 KLGLPQVK-AGRCTNKNPNNTKAPSPALAPALWFGQDGKAEYASVSTYDSSS 368
QY 68 QGDILVGMKKQKE-----RVFPQVNDATSFKTALKTVVPQ-----RITSAAILI-- 112
DB 369 SRIFLQELKQTETGPKGRVSLKSLNDG-----EIKRQPSFNGRQTIIRLDDGVHLIKL 423
QY 113 SDPSQOPLAFVNLGFSNTGL-QALGITDDLGAQPDGQFADANLGGDLDSQWVAPFTGT 171
DB 424 NGSKDEVAAFVNLNGNNTGKNDTFGIKVE-----ANVNLDADEWKVLLPW 469
QY 172 TIHGVELIGSQDDDFLDOFTDDISSTFGSSITQVALSGSARPPDOAGHEHFGFLDGISQ 231
DB 470 TVRG-----PDNDNFKSINOKPEKYSORYIRDNNGNRDLG--DIVNS 511
QY 232 P--SVTGWETTV-----PPG-----QAVVPPGII 253
DB 512 FIVAVGGYLAATAANDGMVHIFKNNGSDERSYNLKLSTYIPGTPMKRDKTQSOESTLAKELR 571
QY 254 LTGRDGDGTGRPSWALDGSFMAFR---HFQKQVPEF-----NAYTLANAIPANSAGN 302
DB 572 AFAEKYGVGDR--YGVDDGFVLRQVELSGQKHVFMFGAMGFGRCGAYAL-----D 619
QY 303 LTQOEGAEFLGARMF-----GRWKSAGPIDLADPTADDPALG----- 338
DB 620 LSKINGNYPAAPLFDVKGODNNGKNRVRKVELGYVTGTPQIGKIRNGKYAFLASGYAAK 679
```

QY 339 -ADPGRNNN-----FDYSDTLTDTRCPFGAHRKTNPRODLCGPVDTFHAMSSIPYGP 393  
DB 680 KIDDSNNKALVYVLDKTL-----GTPIAKIEVXDKG-----LSS-----PT 719  
QY 394 TSDAELASCVT---ADRGLLFVEYOSITGNGFRFOQIN-----MANNANFPPSKITPG 445  
DB 720 LVDPKLDGTVDAVAGDGG-----GNMYRFDLSDSSKMSKVI FEGDKRITSA 769  
QY 446 IEPPIGQTPRTV-----GGP---DPLANQETFTVLFVYIPKG 480  
DB 770 --PAYSRILADKRVIIFGTGSDLTBEDVLTGEGQITVIGIFPDDKG 811

RESULT 11  
US-07-743-357-22  
Sequence 22, Application US/07743357  
Patent No. 5858646  
GENERAL INFORMATION:  
APPLICANT: Kang, Yong C.  
TITLE OF INVENTION: Polypeptide having immunological  
TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: KIRBY EADES GALE BAKER  
STREET: Box 3432, Station D  
CITY: Ottawa  
STATE: Ontario  
COUNTRY: Canada  
ZIP: K1M 1H8  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07743.357  
FILING DATE: 21-AUG-1991  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/CA90/00062  
FILING DATE: 23-FEB-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Gale, Edwin J.  
REGISTRATION NUMBER: 28,584  
REFERENCE/DOCKET NUMBER: 30924-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (613) 237-6900  
TELEFAX: (613) 237-0045  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 913 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
ORGANISM: Human immunodeficiency virus type 1  
STRAIN: HXB2  
US-07-743-357-22

Query Match 3.7%, Score 97.5; DB 2; Length 913;  
Best Local Similarity 22.1%, Pred. No. 1.2;  
Matches 79; Conservative 42; Mismatches 116; Indels 121; Gaps 19;

QY 78 QKERFV---FPOVNDATSKTAL--KTYVPQRITSAILSDPSQO--PLAFVNLGFSNT 130  
DB 494 EKXPVIGATFVVDGAAHNETYLGAGVYVTRGROKVVTLDTTNOXTLQAYLALQDS 553  
QY 131 GLQALGITDD---LGDAPG-PDGFADAAH-LGDDLSQ-----WVAPFTGTTIHGVPL 178  
DB 554 GLEVNIVTDSOVALGIQAGPDSSELVNQLIEQLIKKXVYLAWPARKG-----I 606

QY 179 IGSDDDFLDQFTDDISSFPSSITOVQALSGSARPPDOAGHH-----FGF 225  
DB 607 GGNQVQDKL-----VSNQIRKVLFLDG-----IDKQDHEKXHSNWRMASDFNL 652  
QY 226 LDGTSQPSVYGMETVFPQAV-----VPPGIIITGRDGTGRPSMALDGSFMAFRHQ 280  
DB 653 PPVAKETIVASCDCQLKGEMAHMGQVDCSPGI-----WQLDCTHLEGKYL 698  
QY 281 QKVPENAYTLANAI PANSAGNLPOGAEFLGMRGKRSQAPIDLAPTDAPALGMD 340  
DB 699 VAVHVASGITEAEVLPAR-----TQGETAYFL-LKLAGRM----- 732

QY 341 PGRNNFDYSDTLTDTRCPFGAHRK-----TNPRODLCGPVDTFHAMSSIPYGPET 394  
DB 733 PVTIHTDGSNFT-----GATVRAACWAGIXKQDFG-----ITVYNGS 771

RESULT 12  
US-07-743-357-1  
Sequence 1, Application US/07743357  
Patent No. 5858646  
GENERAL INFORMATION:  
APPLICANT: Kang, Yong C.  
TITLE OF INVENTION: Polypeptide having immunological  
TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: KIRBY EADES GALE BAKER  
STREET: Box 3432, Station D  
CITY: Ottawa  
STATE: Ontario  
COUNTRY: Canada  
ZIP: K1M 1H8  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07743.357  
FILING DATE: 21-AUG-1991  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/CA90/00062  
FILING DATE: 23-FEB-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Gale, Edwin J.  
REGISTRATION NUMBER: 28,584  
REFERENCE/DOCKET NUMBER: 30924-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (613) 237-6900  
TELEFAX: (613) 237-0045  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1005 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
ORGANISM: Human immunodeficiency virus type 1  
STRAIN: HXB2  
US-07-743-357-1

Query Match 3.7%, Score 97.5; DB 2; Length 1005;  
Best Local Similarity 22.1%, Pred. No. 1.4;  
Matches 79; Conservative 42; Mismatches 116; Indels 121; Gaps 19;

QY 78 QKERFV---FPOVNDATSKTAL--KTYVPQRITSAILSDPSQO--PLAFVNLGFSNT 130  
DB 554 GLEVNIVTDSOVALGIQAGPDSSELVNQLIEQLIKKXVYLAWPARKG-----I 606

Db 586 EKEPIVGAETFYVGAANRETKLGAGYVTRGRQKVWLTDTTNTQKTELQAIYIALQDS 645  
 QY 131 GLOALGITDD--LGDQAF-PDGFADAAAN-LGDDLSQ-----WVAPFTGTTIHGVFL 178  
 Db 646 GLEVNIVTDSQYALGIIQAQPDQSESELVQIIEQLIKKEKYVLAVVPAHKG-----I 698  
 QY 179 TOSQDDFLDQFTDDISSTFGSSITVOALSGSARPFDOAGHEH-----PGF 225  
 Db 699 CGNEQVDKL-----VSAGIRKVLFLDG-----IDKAEHEKYHNSNRAMASDNL 744  
 QY 226 LDGISQPSVGTGWTTFVPGQAV-----VPPGIILTGRDGTGTRPSWALDGSFMAFRHQ 280  
 Db 745 PPVAKIEIVASCDKQLKGAMHQVDCSPGI-----WQDCTHLSGKVL 790  
 QY 281 QKVPEFNAYTLANIIPANSAGNLTOQGEAFELGARMFGKWSGAPIDLAPTADDPALGAD 340  
 Db 791 VAVHVASGYIEAEVPAE-----TGQETAYFL-LKLAGRW----- 824  
 QY 341 PORNNNFYSDTLDETRCPFGAHRK-----TNPRDLGPGVDTTFHAMRSSIPYGPET 394  
 Db 825 PVKTIHTDNGSNFT-----GATVRAACWAGIKQBFQ-----IPYNPOS 863

RESULT 13  
 US-07-743-357-9  
 ; Sequence 9, Application US/07743357  
 ; Patent No. 5858646  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kang, Yong C.  
 ; TITLE OF INVENTION: Polypeptide having immunological  
 ; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
 ; NUMBER OF SEQUENCES: 22  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: KIRBY EADES GALE BAKER  
 ; STREET: Box 3432, Station D  
 ; CITY: Ottawa  
 ; STATE: Ontario  
 ; COUNTRY: Canada  
 ; ZIP: K1M 1H8  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/743,357  
 ; FILING DATE: 21-AUG-1991  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/CA90/00062  
 ; FILING DATE: 23-FEB-1990  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Gale, Edwin J.  
 ; REGISTRATION NUMBER: 28,584  
 ; REFERENCE/DOCKET NUMBER: 30924-2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (613) 237-6900  
 ; TELEFAX: (613) 237-0045  
 ; INFORMATION FOR SEQ ID NO: 9:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1003 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: not relevant  
 ; TOPOLOGY: not relevant  
 ; MOLECULE TYPE: protein  
 ; HYPOTHEICAL: NO  
 ; FRAGMENT TYPE: internal  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Human immunodeficiency virus type 1  
 ; STRAIN: MAL  
 ; US-07-743-357-9

Query Match 3.7%; Score 97; DB 2; Length 1003;

Best Local Similarity 21.8%; Pred. No. 1.6;  
 Matches 75; Conservative 43; Mismatches 116; Indels 110; Gaps 18;  
 QY 85 FOVNDATSEFKT--ALKYVVPORITSAAILISDPSQO--PLAFVNLGFSNTGLQALGITDD 140  
 Db 594 FVVDGAANRETKKAGYVTRGRQKVSLTETNTQKTELQAIHLALODSGLEVNIVTDS 653  
 QY 141 ---LGDQAF-PDGFADAAAN-LGDDLSQ-----WVAPFTGTTIHGVFLGSDQDDFLD 188  
 Db 654 QYALGIIQAQPDQSESELVQIIEQLIKKEKYVLAVVPAHKG-----ICGNEQVDKL- 705  
 QY 189 QFTDDISSTFGSSITVOALSGSARPFDOAGHEH-----FGFLDGISOPSVT 235  
 Db 706 -----VSSGIRKVLFLDG-----IDKAEHEKYHNSNRAMASDNLPIVAKEIVA 752  
 QY 236 GWETTVPFGQAV-----VPPGIILTGRDGTGTRPSWALDGSFMAFRHQKVPFNAYT 290  
 Db 753 SCDKQLKGAMHQVDCSPGI-----WQDCTHLSGKIIIVAVHVASGYI 798  
 QY 291 LANATIPANSAGNLTOQGEAFELGARMFGKWSGAPIDLAPTADDPALGADPORNNDYS 350  
 Db 799 EAEVPAE-----TGQETAYFL-LKLAGRW-----PVKVVHT-----DNGSNP--- 835  
 QY 351 DTLDETRCPFGAHRKTNPRDLGPGVDTTFHAMRSSIPYGPET 394  
 Db 836 -----TSAAVKAAACWAGIKQBFQ-----IPYNPOS 861

RESULT 14  
 US-08-485-355B-50  
 ; Sequence 50, Application US/08485355B  
 ; Patent No. 6177075  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Christian, P. D., Gordon, K. H.J., Hanzlik, T. N.  
 ; TITLE OF INVENTION: Insect Viruses and Their Uses in  
 ; Protecting Plants  
 ; NUMBER OF SEQUENCES: 57  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Flehr Hobbach Test Albritton & Herbert LLP  
 ; STREET: Four Embarcadero Center, Suite 3400  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: United States  
 ; ZIP: 94111-4187  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/485,355B  
 ; FILING DATE: 07-Jun-1995  
 ; CLASSIFICATION: <Unknown>  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/440,522  
 ; FILING DATE: 12-MAY-1995  
 ; APPLICATION NUMBER: US 08/089,372  
 ; FILING DATE: 08-JUL-1993  
 ; APPLICATION NUMBER: AU PL4081/92  
 ; FILING DATE: 14-AUG-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Trecartin, Richard P.  
 ; REGISTRATION NUMBER: 31,801  
 ; REFERENCE/DOCKET NUMBER: A-58631-2/RFT/DSS  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 781-1989  
 ; TELEFAX: (415) 398-3249  
 ; TELEX: 910 277299  
 ; INFORMATION FOR SEQ ID NO: 50:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 647 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear

MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 50:  
US-08-485-355B-50

Query Match 3.7%; Score 96; DB 4; Length 647;

Best Local Similarity 21.0%; Pred. No. 0.98;  
Matches 99; Conservative 49; Mismatches 198; Indels 126; Gaps 21;

```

QY 58 NDTILPLNNIGDILVKKKQKRRFVQVNDATSFALKTVPQRTSAAILISDPQ 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184 NDLIELNNLADMRV---VSEQMINF-TNDTT-----YVRRVLRLPTDYDDPTE 232
QY 118 QPLAFVNLGFSNTGIALGITDDIGDAQPDGCFADANIGDDLQGVAPFTGTTIHGVF 177
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 -----GLVRTVSDYRLTYKAITCEANPPLVDD-----GFWIGGY 268
QY 178 LIGSDQDDPLDQFTDDISTFG-SITVOVALSGSARPFQAGHEHFGFLDGIQSPSYTG 236
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 269 ALTPS---LPQY--DVEEAYALHTLTFARPSAAALAFVWAGLPQ---GGTAPAGTFA 319
QY 237 WET-----TVFPGA---VVPFGIILTGKDGDTGRPSMALDGSFMAFRH-- 278
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 320 WEQASSGGYLTWRHNGITTFPAGSVSYVLPFGFALERYPDNDGSMTFASAGDTVTRQVA 379
QY 279 -----FOOKVPEFNAVYTLANAIPANSAGNLTOEGAEFLGAMFGRW 320
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 380 VDEVVYTNPPAGSGSAPFTVTRVPSNAYT--NTVFRNTL--LETPESSRLLELM---- 431
QY 321 KSGAPIDLAFTADDPALGADPQRNNNPDYSDTLTDERCPFGAHVKTNPRLDGGPVD 380
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 432 ---PPADFGQTV-----ANNPKIEQSLKETLGCYLVHSMKRNRPVQL--TPASS 476
QY 381 FHAMRSSIYPGSETSDAELASGVTAAODRGLLFVEYQSIIGNCR-----FOOI 428
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 477 FGVSPNNNGYERTDLDPDTGIRDS-----FDQNSTAVAHRRSLSHSCSIYTKTYQGM 531
QY 429 NNANNANPFPSKPTIPGI---EPI--IGOTTPTVGGDLPLNQNERTYPLF 475
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 532 EGVTVNTPFGQFAHAGLLKNEBILCLADDLATRLTGVPATDNFAAASAF 583

```

RESULT 15  
US-09-194-613-3  
Sequence 3, Application US/09194613  
Patent No. 6251654  
GENERAL INFORMATION:  
APPLICANT: GORDON, Karl H.  
APPLICANT: HANZLIK, Terry N.  
TITLE OF INVENTION: MODIFIED SMALL RNA VIRUSES  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: McDermott, Will & Emery  
STREET: 600 13th Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-1096  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/194,613  
FILING DATE: 30-NOV-1998  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Bucca Ph.D., Daniel  
REGISTRATION NUMBER: 42,368  
REFERENCE/DOCKET NUMBER: 50179-061  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-756-8000  
TELEFAX: 202-756-8087

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 647 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-194-613-3

Query Match 3.7%; Score 96; DB 4; Length 647;

Best Local Similarity 21.0%; Pred. No. 0.98;  
Matches 99; Conservative 49; Mismatches 198; Indels 126; Gaps 21;

```

QY 58 NDTILPLNNIGDILVKKKQKRRFVQVNDATSFALKTVPQRTSAAILISDPQ 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184 NDLIELNNLADMRV---VSEQMINF-TNDTT-----YVRRVLRLPTDYDDPTE 232
QY 118 QPLAFVNLGFSNTGIALGITDDIGDAQPDGCFADANIGDDLQGVAPFTGTTIHGVF 177
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 -----GLVRTVSDYRLTYKAITCEANPPLVDD-----GFWIGGY 268
QY 178 LIGSDQDDPLDQFTDDISTFG-SITVOVALSGSARPFQAGHEHFGFLDGIQSPSYTG 236
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 269 ALTPS---LPQY--DVEEAYALHTLTFARPSAAALAFVWAGLPQ---GGTAPAGTFA 319
QY 237 WET-----TVFPGA---VVPFGIILTGKDGDTGRPSMALDGSFMAFRH-- 278
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 320 WEQASSGGYLTWRHNGITTFPAGSVSYVLPFGFALERYPDNDGSMTFASAGDTVTRQVA 379
QY 279 -----FOOKVPEFNAVYTLANAIPANSAGNLTOEGAEFLGAMFGRW 320
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 380 VDEVVYTNPPAGSGSAPFTVTRVPSNAYT--NTVFRNTL--LETPESSRLLELM---- 431
QY 321 KSGAPIDLAFTADDPALGADPQRNNNPDYSDTLTDERCPFGAHVKTNPRLDGGPVD 380
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 432 ---PPADFGQTV-----ANNPKIEQSLKETLGCYLVHSMKRNRPVQL--TPASS 476
QY 381 FHAMRSSIYPGSETSDAELASGVTAAODRGLLFVEYQSIIGNCR-----FOOI 428
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 477 FGVSPNNNGYERTDLDPDTGIRDS-----FDQNSTAVAHRRSLSHSCSIYTKTYQGM 531
QY 429 NNANNANPFPSKPTIPGI---EPI--IGOTTPTVGGDLPLNQNERTYPLF 475
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 532 EGVTVNTPFGQFAHAGLLKNEBILCLADDLATRLTGVPATDNFAAASAF 583

```

Search completed: November 27, 2002, 13:46:26  
Job time : 27 secs

THIS PAGE BLANK (USPTO)



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 27, 2002, 13:46:05 / Search time 13 Seconds  
(without alignments)  
610.019 Million cell updates/sec

Title: US-09-926-084-7  
Perfect score: 2601  
Sequence: 1 MDLSLFVVAVALVGSSSHV.....KGGEXFLPSPISALTATIAA 498

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 102317 seqs, 15924203 residues

Total number of hits satisfying chosen parameters: 102317

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA:  
1: /cgn2\_6/ptodata/2/pubppa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/2/pubppa/PCPT\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/2/pubppa/US05\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/2/pubppa/US07\_NEW\_PUB pep.\*  
5: /cgn2\_6/ptodata/2/pubppa/US07\_PUBCOMB pep.\*  
6: /cgn2\_6/ptodata/2/pubppa/US07\_PUBCOMB pep.\*  
7: /cgn2\_6/ptodata/2/pubppa/PCPTUS\_PUBCOMB pep.\*  
8: /cgn2\_6/ptodata/2/pubppa/US09\_NEW\_PUB pep.\*  
9: /cgn2\_6/ptodata/2/pubppa/US09\_PUBCOMB pep.\*  
10: /cgn2\_6/ptodata/2/pubppa/US10\_NEW\_PUB pep.\*  
11: /cgn2\_6/ptodata/2/pubppa/US10\_PUBCOMB pep.\*  
12: /cgn2\_6/ptodata/2/pubppa/US60\_NEW\_PUB pep.\*  
13: /cgn2\_6/ptodata/2/pubppa/US60\_PUBCOMB pep.\*  
14: /cgn2\_6/ptodata/2/pubppa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	137.5	5.3	439	US-09-815-242-11712	Sequence 11712, A
2	130.5	5.0	423	US-09-815-242-10121	Sequence 10121, A
3	118	4.5	2383	US-09-812-020-302	Sequence 302, App
4	105	4.0	753	US-09-815-242-10181	Sequence 10181, A
5	101	3.9	398	US-09-815-242-15469	Sequence 5469, Ap
6	101	3.9	409	US-09-815-242-12418	Sequence 12418, A
7	96	3.7	1349	US-09-815-242-5898	Sequence 5898, Ap
8	96	3.7	1349	US-09-815-242-13137	Sequence 13137, A
9	95	3.7	991	US-09-815-242-5803	Sequence 5803, Ap
10	92.5	3.6	978	US-10-007-693-65	Sequence 65, Appl
11	92	3.5	349	US-09-819-935-2	Sequence 2, Appl
12	91	3.5	707	US-09-764-870-313	Sequence 313, App
13	90	3.5	637	US-09-852-053-3	Sequence 3, Appl
14	90	3.5	931	US-09-779-081-2	Sequence 2, Appl
15	90	3.5	1609	US-10-007-693-75	Sequence 75, Appl
16	89.5	3.4	646	US-09-777-710A-15	Sequence 1, Appl
17	89.5	3.4	670	US-09-777-710A-15	Sequence 15, Appl
18	89.5	3.4	3472	US-10-027-806-4	Sequence 4, Appl
19	88	3.4	376	US-09-815-242-10160	Sequence 10160, A

20	88	3.4	870	US-09-815-242-5493	Sequence 5493, Ap
21	88	3.4	870	US-09-815-242-12637	Sequence 12637, A
22	87.5	3.4	628	US-09-815-626-2	Sequence 2, Appl
23	87.5	3.4	5215	US-09-860-846-2	Sequence 2, Appl
24	87.5	3.4	5215	US-09-861-289-2	Sequence 2, Appl
25	87	3.3	529	US-09-731-393-12	Sequence 12, Appl
26	87	3.3	1047	US-09-866-562-57	Sequence 57, Appl
27	86	3.3	334	US-09-815-242-13249	Sequence 13249, A
28	86	3.3	572	US-09-942-185-4	Sequence 4, Appl
29	86	3.3	1477	US-10-092-880-4	Sequence 4, Appl
30	85	3.3	516	US-09-731-393-17	Sequence 17, Appl
31	84	3.2	333	US-09-897-898-3	Sequence 3, Appl
32	84	3.2	333	US-09-897-898-14	Sequence 4, Appl
33	84	3.2	333	US-09-897-898-14	Sequence 14, Appl
34	84	3.2	388	US-09-897-898-17	Sequence 17, Appl
35	84	3.2	451	US-09-897-898-19	Sequence 19, Appl
36	84	3.2	568	US-10-080-233-4	Sequence 4, Appl
37	84	3.2	1537	US-09-803-368-104	Sequence 104, App
38	83.5	3.2	716	US-10-036-492-11	Sequence 11, Appl
39	83.5	3.2	722	US-09-853-533A-10	Sequence 10, Appl
40	83	3.2	660	US-09-841-132-446	Sequence 446, App
41	83	3.2	887	US-09-815-242-13179	Sequence 13179, A
42	83	3.2	996	US-08-910-386A-5	Sequence 5, Appl
43	83	3.2	6281	US-09-815-242-12996	Sequence 12996, A
44	82.5	3.2	443	US-10-083-452-5	Sequence 5, Appl
45	82.5	3.2	611	US-09-829-549A-48	Sequence 48, Appl

#### ALIGNMENTS

##### RESULT 1

Sequence 11712, Application US/09815242

Patent No. US20020061859A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl L.

APPLICANT: Zykkind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815.242

PRIOR FILING DATE: 2001-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

PRIOR FILING DATE: 2000-03-21

Query Match 5.3% Score 137.5; DB 10; Length 439;  
Best Local Similarity 21.2%; Pred. No. 7.9e-05;





```
; PRIOR APPLICATION NUMBER: 60/206,848
; CURRENT FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5469
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5469
```

```
Query Match          3.9%; Score 101; DB 10; Length 398;
Best Local Similarity 20.3%; Pred. No. 0.15;
Matches 95; Conservative 62; Mismatches 158; Indels 154; Gaps 23;
```

```
QY 42 QAPLPTLQHTTESGANDTILPLNNIOGDILVGMKKQKERVFFQVNDATSKFTALKTYV 101
DB 53 QPGITPTQKT-----CNFVALDLKSKDRDAIKAMFKKWTVMADRMMDGDTVGKTSNNPLM 108

QY 102 PORITSAAILISDPSQOPLAFVNLGFSNTGLQALGITDDLGA-----QFPGQFADAAN 156
DB 109 PVDVTGESIGIG-ASKLITITF---GISKLMKKIGLSSKIPDAFKDLPHFPNDQLIDYIS 164

QY 157 LGDDLSQWVAFTTTHGVFLIGSDODDLDFTDDISSTFGSSITQVQALSGSARPPD 216
DB 165 DGDIMIQAACNSDSQVSFAV-----HNLVRPFRDIVKVRWSQS-----202

QY 217 QAGHEHFGFLDGIQSPSVGTWETVFPQAVVPPGIIILTRGDGTGTRPS-----266
DB 203 -----GFI-----SAGKET-----PRNLMAFKDGTINPRKSNQLKDYVPID 239

QY 267 --WALDGSFMAFRHQKVPFENAYTLANAIPANSAGNLTQOEGAEFLGARFGRWK-SG 323
DB 240 DGMWAKHGYCVVRRIQIHETWDRTA-----LEEQEA-----FGRKRHS 280

QY 324 APIDLAPTADDPALGADPQRNNFDYSDTLTDETRCPFGAHRKTNPRDLGGPVDTFHA 383
DB 281 APLTGKGFEDIDLKA---KDSHGEY---IIDK-----DAHTR-----LAKEANT-SI 321

QY 384 MRSSIPYGPETSDEALASGVTAQ-DRGLLVEYQSIIGNGFRRQOINWANNANFFPSKPI 442
DB 322 LRARFNVYDGTDDR-----TCGNFETGLLFIATQKATQ-----QFIDIQNN-----361

QY 443 TPGIEPIIGTTPRTVGGDLPLNQNFTVPLFVIPKGEYFF-LPSIS 490
DB 362 -----LGSNDKLE-----YITHRGSASFVLVPGVS 387
```

```
RESULT 6
US-09-815-242-12418
; Sequence 12418, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zykand, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; PROKARYOTES
; FILE REFERENCE: ELITRA.011a
```

```
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12418
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12418
```

```
Query Match          3.9%; Score 101; DB 10; Length 409;
Best Local Similarity 20.3%; Pred. No. 0.15;
Matches 95; Conservative 62; Mismatches 158; Indels 154; Gaps 23;
```

```
QY 42 QAPLPTLQHTTESGANDTILPLNNIOGDILVGMKKQKERVFFQVNDATSKFTALKTYV 101
DB 64 QPGITPTQKT-----CNFVALDLKSKDRDAIKAMFKKWTVMADRMMDGDTVGKTSNNPLM 119

QY 102 PORITSAAILISDPSQOPLAFVNLGFSNTGLQALGITDDLGA-----QFPGQFADAAN 156
DB 120 PVDVTGESIGIG-ASKLITITF---GISKLMKKIGLSSKIPDAFKDLPHFPNDQLIDYIS 175

QY 157 LGDDLSQWVAFTTTHGVFLIGSDODDLDFTDDISSTFGSSITQVQALSGSARPPD 216
DB 176 DGDIMIQAACNSDSQVSFAV-----HNLVRPFRDIVKVRWSQS-----213

QY 217 QAGHEHFGFLDGIQSPSVGTWETVFPQAVVPPGIIILTRGDGTGTRPS-----266
DB 214 -----GFI-----SAGKET-----PRNLMAFKDGTINPRKSNQLKDYVPID 250

QY 267 --WALDGSFMAFRHQKVPFENAYTLANAIPANSAGNLTQOEGAEFLGARFGRWK-SG 323
DB 251 DGMWAKHGYCVVRRIQIHETWDRTA-----LEEQEA-----FGRKRHS 291

QY 324 APIDLAPTADDPALGADPQRNNFDYSDTLTDETRCPFGAHRKTNPRDLGGPVDTFHA 383
DB 292 APLTGKGFEDIDLKA---KDSHGEY---IIDK-----DAHTR-----LAKEANT-SI 332

QY 384 MRSSIPYGPETSDEALASGVTAQ-DRGLLVEYQSIIGNGFRRQOINWANNANFFPSKPI 442
DB 333 LRARFNVYDGTDDR-----TCGNFETGLLFIATQKATQ-----QFIDIQNN-----372

QY 443 TPGIEPIIGTTPRTVGGDLPLNQNFTVPLFVIPKGEYFF-LPSIS 490
DB 373 -----LGSNDKLE-----YITHRGSASFVLVPGVS 398
```

```
RESULT 7
US-09-815-242-5898
; Sequence 5898, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zykand, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
```

APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-28  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 1410  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5898  
LENGTH: 1349  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-5898

Query Match 3.7%; Score 96; DB 10; Length 1349;  
Best Local Similarity 20.9%; Pred. No. 2.5;  
Matches 57; Conservative 34; Mismatches 100; Indels 82; Gaps 12;

QY 37 TNPQGAAPLITQHTTSGANDTILPLNINOGLVGMKKKE-RVFFQVNDATSEFT 95  
DB 803 TNDGK-----QSTKSGISGVTTLKENGVLQTTTDXDKXQFTGLENGT-YKV 854  
QY 96 ALKTT--YVPRITSAAILISDPGQPLAFVNLGFSNTGLQALGITDDIGDAQFPDQPA 152  
DB 855 EFTPGYPTQVSG-----TDEGIDSGNSTTGVIMDKNDITDGGFYK 900  
QY 153 DAANIGDDLSQW-----VAPFTGTHGVFLIGSODPFL-----DQFTDOI 194  
DB 901 PTVNLGDVY--WEDTNKNGVQDKDEKISGVTTLKENDVLTCTTTDENGKYQFTDLN 958  
QY 195 SSTFGSSITQVQALSGSARPPDQAGHEHFGFLDGISQPSVTGMETTVFPGQAVVPPGIL 254  
DB 959 NGTY-----KVEFETPSGYTPTSVTSNDTEKDSNGLTGTVI- 996  
QY 255 TGRDQDTGRPSMALDGSFMAFRHQVPEFN 287  
DB 997 --KDADNMT-----LDGPF-----YKTPKYS 1015

RESULT 8  
US-09-815-242-13137  
Sequence 13137, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 1410  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13137  
LENGTH: 1349  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-13137

Query Match 3.7%; Score 96; DB 10; Length 1349;  
Best Local Similarity 20.9%; Pred. No. 2.5;  
Matches 57; Conservative 34; Mismatches 100; Indels 82; Gaps 12;

QY 37 TNPQGAAPLITQHTTSGANDTILPLNINOGLVGMKKKE-RVFFQVNDATSEFT 95  
DB 803 TNDGK-----QSTKSGISGVTTLKENGVLQTTTDXDKXQFTGLENGT-YKV 854  
QY 96 ALKTT--YVPRITSAAILISDPGQPLAFVNLGFSNTGLQALGITDDIGDAQFPDQPA 152  
DB 855 EFTPGYPTQVSG-----TDEGIDSGNSTTGVIMDKNDITDGGFYK 900  
QY 153 DAANIGDDLSQW-----VAPFTGTHGVFLIGSODPFL-----DQFTDOI 194  
DB 901 PTVNLGDVY--WEDTNKNGVQDKDEKISGVTTLKENDVLTCTTTDENGKYQFTDLN 958  
QY 195 SSTFGSSITQVQALSGSARPPDQAGHEHFGFLDGISQPSVTGMETTVFPGQAVVPPGIL 254  
DB 959 NGTY-----KVEFETPSGYTPTSVTSNDTEKDSNGLTGTVI- 996  
QY 255 TGRDQDTGRPSMALDGSFMAFRHQVPEFN 287  
DB 997 --KDADNMT-----LDGPF-----YKTPKYS 1015

RESULT 9  
US-09-815-242-5803  
Sequence 5803, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27

;; PRIOR APPLICATION NUMBER: 60/257,931  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/269,308  
;; PRIOR FILING DATE: 2001-02-16  
;; NUMBER OF SEQ ID NOS: 14110  
;; SOFTWARE: fastseq for Windows Version 4.0  
;; SEQ ID NO 5803  
;; LENGTH: 991  
;; TYPE: PRT  
;; ORGANISM: Staphylococcus aureus  
US-09-815-242-5803

Query Match 3.7%; Score 95; DB 10; Length 991;  
Best Local Similarity 22.1%; Pred. No. 2;  
Matches 98; Conservative 48; Mismatches 184; Indels 114; Gaps 19;

QY 7 VYVAVLVGSSSHVNAKLGAROTRTTPLL--TNFFG---QAPLPTLTQHTTESGANDTI 61  
DB 145 VASKLTDGTGSSVVVKEANEQFNKTVTRALLEEANKAGLTIEENVPTINKITNAGYSADKA 204  
QY 52 LP-----LNNIQ-----GDILVGMKKOKE-RFVFFQVNDATS 92  
DB 205 LPKINDFANRIVYUNNHQADLDKYASDFRQLGSGYEGDILDASEKIKRINGAIPQUNEKAK 264  
QY 93 FKTALKTYVQRTISAAILISD--PSQOPLAFVNLGFSNTGL-QALGITDD-----LGDAQ 145  
DB 265 LIALANNYMP-KIEKALNFAADDVPAQFPKINQGLNIASOGIDQANGQLNDKAGFVTVQR 323  
QY 146 FPDQGFADAANLGDGL--SQWAPFTGTHGVFLIGSDODDFLDODTDDISSTFFG--- 199  
DB 324 SRVGDYQAIRRAQDLNRRQOQIP-----QNSAANNETSNSAPAAGNGV 368  
QY 200 -----SSITQVQALSGSARFPDQAGHEHFGFLDGISQPSVTGWTTPPG 244  
DB 369 ASTPSPASGDTAPNNNVNTQNTAPNSNNAPVSTTPQSTGKKGQSFVDIT---TT----- 421  
QY 245 QAVVPPGIIILTRDGDGTGTRPSWALDGSFNAFRHFOOKVPEFNAYTLANAIPANSAGNLT 304  
DB 422 QVSTANENTQNTDKVKSEEA-ALTGSLLSLN-----NLDTQAKA----- 462  
QY 305 QOEGAEFLGARMFGRWKSGAPIDLAPTADDPALGADPORNNNFDYSIDLTDTRCPFGAH 364  
DB 463 AQKQSOALRNISYGLASDKPSPFRESLDNVKSGLEVYTYQYNOQFDITLKE----- 513  
QY 365 VRKNTNPRODGGPVDTTHMRSSI 388  
DB 514 IEK-NEENVDLKSIEDKVKAAANNRI 536

RESULT 10  
US-10-007-693-65  
;; Sequence 65, Application US/10007693  
;; Patent No. US20020146776A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Bhactia, Ajay  
;; APPLICANT: Probel, Peter  
;; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT  
;; TITLE OF INVENTION: AND DIAGNOSIS OF CHLAMYDIAL INFECTION  
;; FILE REFERENCE: 210121.515G2  
;; CURRENT APPLICATION NUMBER: US/10/007,693  
;; CURRENT FILING DATE: 2001-12-05  
;; NUMBER OF SEQ ID NOS: 157  
;; SEQ ID NO 65  
;; LENGTH: 978  
;; TYPE: PRT  
;; ORGANISM: Chlamydia pneumoniae  
US-10-007-693-65

Query Match 3.6%; Score 92.5; DB 12; Length 978;  
Best Local Similarity 20.8%; Pred. No. 3.2;  
Matches 105; Conservative 47; Mismatches 213; Indels 141; Gaps 23;  
QY 9 SVAVLVGSSSHVNAKLGAROTRTTPLLTFNFPQAPLPTLTQHTTESGANDTILPLNNIQ 68

DB 396 AITVEAGGEIVLSAOGSRLVFYDPIHSLP-----TTSPSNKOITINANGAS 444  
QY 69 GDILVGMKKOKEKRFVFFQVNDATSFKTALKTYVQRTISAAILISDPSQOPLAFVN-LGF 127  
DB 445 GSVVFTSKGLSSTELLIPANTTTILLGTVK-----IASGELKITDN-----AVNVVLGF 493  
QY 128 SNTGLQALGITDGLGDAQFPDQGFADAANLGDGLSOWAPFTGTHGVFLIGS---DOD 184  
DB 494 ATQG-----SGQL-----TLGSGGTGLGATPTGAPAAVDFITIGKLAPDPF 533  
QY 185 DFLQDFTDDISSTFGSSITQVQALSGSARFPDQAGHEHFGFLDGISQPSVTGWTTPPG 244  
DB 534 SFLXR--DFVSASVNACTKNV-TLTG-ALVLDE--HDVTDLYDMVLSQSPVAIPAFVK 587  
QY 245 QAVVPPGIIILTRDGDGTGTRPSWALDGSFNAFRHFOOKVPEFNAYTLAN--AIPANSAGN 302  
DB 588 ATVTKTGP---PDGEIATPSHYGYQGW-----SYTWSRPLLIAPADGG- 628  
QY 303 LTQOEGAEFLGA-----RMFGRWKSAPIDLAPTADDPALGADPORN-----NNFDYS 350  
DB 629 -----FPGGSPSANTLYAVNSDTLVRSTYILDERYGEIVNSLWISFLGNQAFS 680  
QY 351 DTLTDE-----TRCPFGAHVRKTNPRQDLGGPVDTTHMRSSIPYGPETSDELA 400  
DB 681 DILQDVLILLHPGLSITAKALGAYVEHT-PROGHEG-----F 716  
QY 401 SGVTAQDRGLLFEVYQSIIGNGFRFOQINWANNANFPFSKPIPT-----GIEPIIGQ 452  
DB 717 SORVGGYQAALSMMYTDHTTLGLSFGQLYKTNAN-PYDSRCSEQMYLLSFPQFPVITQ 775  
QY 453 TTPRTVG-----GLDPLNQNETHVTP 473  
DB 776 KSEALISWKAAYGYSKNHLNTTYLRP 801

RESULT 11  
US-09-919-935-2  
;; Sequence 2, Application US/09919935  
;; Patent No. US20020049305A1  
;; GENERAL INFORMATION:  
;; APPLICANT: BATHE, BRIGITTE  
;; APPLICANT: MOCKEL, BETTINA  
;; APPLICANT: PFEFFERLE, WALTER  
;; APPLICANT: HUTHMACHER, KLAUS  
;; APPLICANT: BINDER, MICHAEL  
;; APPLICANT: GREISSINGER, DIETER  
;; APPLICANT: THIERBACH, GEORG  
;; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES WHICH CODE FOR THE METF GENE  
;; FILE REFERENCE: 211712USOX  
;; CURRENT APPLICATION NUMBER: US/09/919,935  
;; CURRENT FILING DATE: 2001-08-02  
;; PRIOR APPLICATION NUMBER: DE 10053942.4  
;; PRIOR FILING DATE: 2000-08-02  
;; PRIOR APPLICATION NUMBER: DE 10109686.0  
;; PRIOR FILING DATE: 2001-02-28  
;; PRIOR APPLICATION NUMBER: US 60/294,279  
;; PRIOR FILING DATE: 2001-05-31  
;; NUMBER OF SEQ ID NOS: 4  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 2  
;; LENGTH: 349  
;; TYPE: PRT  
;; ORGANISM: Corynebacterium glutamicum  
US-09-919-935-2

Query Match 3.5%; Score 92; DB 10; Length 349;  
Best Local Similarity 23.0%; Pred. No. 0.82;  
Matches 45; Conservative 24; Mismatches 73; Indels 54; Gaps 7;  
QY 2 DLSLFVSVAVLVGSSSHVNAKLGAROTRTTPLLTFNFPQAPLPTLTQHTTESGANDTI 61  
DB 76 DLGASFVSVTVGAGGSTRERTSRI-ARRLAKOPLTT-----LVHLLTVNHTRE----- 122

```
Qy 62 LPUNNIGDILVGMKKKQKRFVFPQVNDATSPKTAALKTYVPORITSAAILISDPQOPLA 121
Db 123 -----EMKALIKRYELGLITNLLALRGPPGDDPLG 152
Qy 122 -FVN---LGFNTGIALGITD-----DLGDAOPDGOFPADANLGDLSQWVAPFTGT 171
Db 153 DMVSTGGAGNVASELIDLIKSTPREFRPDLGIAFPEGHFR-AKLTBEDTKYTLAKLRG 211
Qy 172 THGVFLIGSDDDPL 187
Db 212 AEYSITQMFEDVEDYL 227

RESULT 12
US-09-764-870-313
; Sequence 313, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT214
; CURRENT APPLICATION NUMBER: US/09/764, 870
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 313
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-870-313

Query Match 3.5%; Score 91; DB 10; Length 707;
Best Local Similarity 20.3%; Pred. No. 2.8; Indels 172; Gaps 26;
Matches 96; Conservative 63; Mismatches 143;

Qy 9 SVAVLVSSSHVNAAKLGAQRTPTPLTNPGQAPLPTLTQHTTESGANDTILPLNNIQ 68
Db 237 SVADVVEASHIE--ORGCTETRTA-----FGIEFARSEAFQKGRGAKKVMIVITDGS 289
Qy 69 GDILVGMKK--OKERFVFQVNDATSEKTAALKTYVPOR-----ITSAAILISDPQO 118
Db 290 SHDSPDEKXTQOSER-----DNVTRAVAVLGYNNRGINPTFLEIKTYIASDDDK 343
Qy 119 PLAFVNLGFSNTGLQAL-GITDDLDAQFP--DGFADANLGDLSQWVAPFTGTTIH-- 174
Db 344 HF-----FNVTEALAKDIDALGDRIFSLGTTKQKTSFGLMSQ-----TGFSSHV 392
Qy 175 --GVFL--IGSD-----QDDFLDQTTDISSTFGSSITVOVALSGA 212
Db 393 EDDVLLGAVGANVDMGAVIKETSAGKVIPLRESYLKEPBEELKN----- 436
Qy 213 RPFDAQHEHFGFLDGISQPSVTGWETTFPQCAVVPFGIILTGRDGTGRPSMALDGS 272
Db 437 -----HGAVL-----GYTVT-----SVMSRQG----- 454
Qy 273 FMAFRHQOKVPEFN-----AYTLANALPANSAGNLTOOGAFLGARFGRKSGAP 326
Db 455 ---RYVVGAPRFNHTKVILFTWHN-----NRSLLTHQAMRGQIGS-YFGSEIRTSVD 505
Qy 327 DLAPYADDPALGAD-----PQNNNFYDSDLTDE---TRCPFG---AHV 365
Db 506 DGGGVTVLLVGAPEVFNENGRGRKVVYELRQNLFFVNGTLKXSHSYQVARFOSSIASV 565
Qy 366 RKNPNPQD-----LGGPVD-----TFHAMRSSITYGPE--TSDAELASGV 403
Db 566 KDLN--ODSINDVAVGAPLEDNHAIAIYFHFGRGSLIKTKRKORITASELATGL 617

RESULT 13
US-09-852-053-3
; Sequence 3, Application US/09852053
```

```
; Patent No. US20020055141A1
; GENERAL INFORMATION:
; APPLICANT: BERENS, STEPHAN
; APPLICANT: KALINOWSKI, JORN
; APPLICANT: PUHLER, ALFRED
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM STRAIN WITH
; FILE REFERENCE: MAS/21123/280248
; CURRENT APPLICATION NUMBER: US/09/852, 053
; CURRENT FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: EPO 00110021.3
; PRIOR FILING DATE: 2000-05-12
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; NAME/KEY: PROPEP
; LOCATION: (1)..(637)
; OTHER INFORMATION: secD
US-09-852-053-3

Query Match 3.5%; Score 90; DB 10; Length 637;
Best Local Similarity 18.6%; Pred. No. 3;
Matches 72; Conservative 62; Mismatches 118; Indels 136; Gaps 19;

Qy 9 SVAVLVSSSHVNAAKLGAQRTPTPLTNPGQAPLPTLTQHTTESGANDTILPLNNIQ 68
Db 97 TLVITVPCENTFAOASLG-----QTSOLLFRVCGAGMPDMT-----TLMP----- 137
Qy 69 GDILVGMKKQKRFVFP-----QVNDATSEKTAALKTYVPORITSAAILI 112
Db 138 -----ELEEMANRWIEGYITTEBOANASLEEMNTRAVASTTAVGEEETEPYTSATW 192
Qy 113 SDBSQOPLAFVNLGFSNTGLQALGITDDLDAQFPDQFAD-----AANLGDLSQWVA 166
Db 193 DEBA-----NSIEATQRCQETIDMLRTDRQSTDPVQIAASSLMQCTTDEWD 239
Qy 167 PFTGT-----THGVFLIGSDDDFLDQFTDDISSR--RQSSITVOVALSGA 212
Db 240 PLAGTDDPRLPLVACDPAGGVVVL--DPAPLNGETDEEGARLTGENIDTNRPITGCP 297
Qy 213 RPFDAQHEH--FGFLDGISQPSVTGWETTFPQCAVVPFGIILTGRDGTGRPSMALD 270
Db 298 NA--QSGQWEISFAPKSG-----DGEES-ATW----- 322
Qy 271 GSFMAFRHQOKV-----PEFNATTLANALPANS-----GNLTOOGAFLGARFGRM 320
Db 323 -SSLSQVLDQOAIATLDSQVISAIVQSAIVPVSATVSGITGDTTEADQDLANNLRVG-- 379
Qy 321 KSGAPIDLA-----PTADDPALGA 339
Db 380 --ALPLSPFAGENGERRGTTTVPPELGA 405

RESULT 14
US-09-779-081-2
; Sequence 2, Application US/09779081
; Patent No. US20020102270A1
; GENERAL INFORMATION:
; APPLICANT: Murrin, Andrew
; TITLE OF INVENTION: CHLAMYDIA ANTIGENS AND CORRESPONDING DNA FRAGMENTS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 19721-009
; CURRENT APPLICATION NUMBER: US/09/779, 081
; CURRENT FILING DATE: 2001-02-08
; PRIOR APPLICATION NUMBER: 09/427, 533
; PRIOR FILING DATE: 1999-10-26
; PRIOR APPLICATION NUMBER: 60/106, 046
; PRIOR FILING DATE: 1998-10-28
; PRIOR APPLICATION NUMBER: 60/132, 271
```





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2002, 14:49:27 ; Search time 2820 Seconds

(without alignments)  
15418.295 Million cell updates/sec

Title: US-09-926-084-8

Sequence: 1 atgcgcctgcgcctgttgc.....tcactgcactacgcctgct 1494

Scoring table: IDENTITY NUC

Gapop 10'-0, Gapex 1 10

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_hcg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*  
29: em\_vl: \*  
30: em\_hcg\_hum: \*  
31: em\_hcg\_inv: \*  
32: em\_hcg\_other: \*  
33: em\_hcg\_mus: \*  
34: em\_hcg\_pln: \*  
35: em\_hcg\_rtd: \*  
36: em\_hcg\_mam: \*  
37: em\_hcg\_vrt: \*  
38: em\_sy: \*  
39: em\_hngo\_hum: \*  
40: em\_hngo\_mus: \*  
41: em\_hngo\_other: \*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1494	100.0	1494	6	E55173
2	1494	100.0	1586	8	AB013135
3	399.8	26.8	1798	8	PSU77073
4	325.6	21.8	1735	8	AF468656
5	62.4	4.2	125020	9	AF429315
6	54.6	3.7	125020	9	AF429315
7	54.6	3.7	213732	1	AE001862
8	52	3.5	2040	1	AY034474
9	51.8	3.5	34713	1	SCF42
10	46.2	3.1	2150	3	DROACT88F
11	45	3.0	28527	1	SCB68
12	44.4	3.0	20721	1	SC3C9
13	44	2.9	11738	1	AE010452
14	43.4	2.9	20235	1	SERERYAB
15	43.4	2.9	20235	6	AR049368
16	43.4	2.9	20235	6	AR095529
17	43	2.9	1136	3	DROACT8F
18	43	2.9	4755	3	AB003910
19	43	2.9	14866	1	SC1B5
20	42.4	2.8	3454	8	AF432502
21	42.4	2.8	3468	8	AF432503
22	42.4	2.8	9755	14	RUBCG
23	42.4	2.8	9759	6	163470
24	42.4	2.8	9762	6	AX009468
25	42.4	2.8	9762	14	AB047329
26	42.4	2.8	9762	14	AB047330
27	42.4	2.8	9762	14	AF188704
28	42.4	2.8	141293	2	AP004337
29	42.4	2.8	142418	2	AP005126
30	42.4	2.8	151008	2	AP004261
31	42	2.8	101546	2	AP004004
32	42	2.8	167114	2	AP004685
33	41.8	2.8	15046	1	AE007176
34	41.8	2.8	121125	1	MTV025
35	41.6	2.8	10299	1	AE004585
36	41.6	2.8	33517	1	SC10B7
37	41.4	2.8	8358	1	DROFLIAC
38	41.4	2.8	8296	2	AC014161
39	41.4	2.8	167918	3	AC007769
40	41.4	2.8	223517	3	AE003709
41	41.2	2.8	861	6	AX458796
42	41.2	2.8	861	6	AX458798
43	41.2	2.8	861	6	AX459800
44	41.2	2.8	861	6	AX459806
45	41.2	2.8	861	6	AX459808

#### ALIGNMENTS

RESULT 1  
E55173  
LOCUS E55173  
DEFINITION Novel enzyme having decoloring activity and method for decoloring  
dye by using the same.  
ACCESSION E55173  
VERSION E55173.1 GI:18629763  
KEYWORDS JP2000245468-A/1.  
SOURCE unclassified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 1494)  
AUTHORS Shoda,M., Kanno,Y. and Kubota,H.  
TITLE Novel enzyme having decoloring activity and method for decoloring  
dye by using the same

Pred. No. is the number of results predicted by chance to have a





U77073.1 GI:2160704  
polyporaceae sp.  
Polyporaceae sp.  
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
Aphyllophorales; Polyporaceae.  
Han, Y.-H.  
1 (bases 1 to 1798)  
Direct Submission  
TITLE  
Submitted (04-NOV-1996) Microbiology, Seoul National University,  
San 56-1, Shillim-dong, Kwanak-gu, Seoul 151-742, Korea  
FEATURES  
Source  
1..1798  
/organism="Polyporaceae sp."  
/db\_xref="taxon:5443"  
1..1798  
/gene="cpop21"  
1..1500  
/gene="cpop21"  
/notes="oxidizes phenolic lignin model compounds"  
/codon\_start=1  
/product="peroxidase"  
/protein\_id="AAB58908.1"  
/db\_xref="GI:2160705"  
TRANSLATION="MQLKHLAATAAFSAVTSAPAYHVKRARSPLIGSFPOQPPLP  
TIAEVQSSAGNDSLPFENIOGDIIVGMKKDKKEFVFFHINNATFRSFLKYAPANI  
TSVOTITGPASGQPAFVNLFASFHFGALGVADLDQDTAFTAGQFADAPSGAEKGT  
WEAFKGTNDGVFLIGSDVITTYOYRDLLKAKLGDAMTVLLDLSAARPGAEKHE  
HRVLDIGSNPTIPGFTGPGQAVVDGVI FAGFTNDPVTNRPSWALDSFLVFRKL  
KQLVPEHKWTLNALONQAGNLTVBEGALLGSRMFGRNWAGPIBLTDPDQDPTLG  
NDPORNDFNYIHPGDELITDETRCPPTAHVRKTNPROLESAOGLIPOLFAIRAGTPY  
GPEVTDAESNNTSIDRLGFLAFVQSVISNGFRFQQLNWMANNAPFNKSEPLGLDP  
VIGQGTGRLTGLDPRNASLPTQIISNGEYFSPSITALVEHFGA"  
BASE COUNT 420 a 487 c 423 g 468 t  
ORIGIN  
Query Match 26.8%; Score 399.8; DB 8; Length 1798;  
Best Local Similarity 58.8%; Pred. No. 9e-79;  
Matches 834; Conservative 0; Mismatches 552; Indels 33; Gaps 7;  
QY 79 GCGAGACAGCGGTACGACACCCCTCTCACTAAATTTCCGGGACAAGCCCCGTGCGG 138  
DB 73 GTGAAGCGGGAAGATCGTCTCCCTCATTTGGGTCTATTCCTCGGTGAGCCACTCCCT 132  
QY 139 ACTCTAACCGACGATACGACGAGCGGGCCAAAGATACATTTCTGCGGCGATCAAC 198  
DB 133 ACTATCGGAGGTTTCACTCTCTCTCGGGGAAATGATAGCTTCCGTTGAGATA-- 190  
QY 199 ATACAAGCGGACATTTGGTTGGCATGAAGAAACAGAGAAACGCTTCCTTTTCCAA 258  
DB 191 -TCCAGGGAGATATTTCTGTGGCATGAAGAAGGACAAAGAGAAGTTTGTCTTCTCCAC 249  
QY 259 GTCAATGACGAACTGTTCAACGCGCTTGAAGACCTAGCGCTGAGCGCATCAG 318  
DB 250 ATTAACAATGACACAGCTTCAAGAGTTTCTTGAAGACGATGCTCCCGCAATATCACT 309  
QY 319 TCGGCGCGCATTTTGAATTCAGATCTCTCTCAGACCGGTTCCTTCCTCAACCTCGGG 378  
DB 310 TCTGTCCAAAACATCATCGGTCTGCTCGGTTCAGCCCCAGCGGTTTGTGAACCTTCGG 369  
QY 379 TTTTTCGAACACAGCGCTCCAGCGCTTGGAAATACCGACGATCTGGGTGATGCACAAATTC 438  
DB 370 TTCTCCCATACCGGTTTCGGTGTCTGGGTGTTGCCAATGATTTACAGATACTCCTTTC 429  
QY 439 CCAGATGTGTAGTTTCGACAGCGCGCAAACTCCGGGACGACCTCAGCCAAATGGGTGCG 498  
DB 430 ACAGCAGGACAGTTTGGCGAGCGCCCTTCTTTTGGCGACACACAGCAGACTTTGGGAGAA 489  
QY 499 CTTTTTACTGTGTACACCATCATGTGTCTTTCTGATGGTAGCGACACGACGACTTC 558  
DB 490 GCTTTTCAAGGGACCAACGTTGATGGGGTTTTCTTGAATGGTAGCGCATGTCAACCACC 549  
QY 559 TTGGATCATGTTACGGATGATATCTCTTCGACCTTTGGTTCTCTCCATCACTCAGGTGCG 618

Db 550 ACAAAACCAATACCGGATGACCTCAAGGCTAAGCTCGGTGAGCGTTGGACGCTCTTGCTC 609  
QY 619 GGGCTCAGTGGGTCTGCGCGTCCAGGAGATCAGGCTGGTTCATGAACACTTTGGGGTTCTCTC 678  
Db 610 GATCTTGACATGCAGCAAGCGCGGGCTGAGAGGGACACGACATTTTGGCTACTTGG 669  
QY 679 GACGCATCTCGACGCCCTCAGTCACAGGCTGGGAGACACGCTTTCCTTGGACAGCGG 738  
Db 670 GACGATATCTCAATCTCTACTATCCCGGCTTCGGTACACC--ATTCCCTGGCCAAGCC 726  
QY 739 GTCGTCCTCCACCTGGAATATCTCTACCTGACGCGATGGGACACAGGGAC--CCGACCG 795  
Db 727 GTAGTGGATTCCGGTGTATCTTTCGCGCGCGCACCAAGTCCCTGACCAACGCTCT 786  
QY 796 TCGTGGCTCTAGATGGGAGTTTTCATGGCAATTCGGGCACTTCCAGCAGAGAGTTCGCCGAA 855  
Db 787 TCGTGGCTCTGGATGGTCTCTTTCTTTTTCGCAAGCTCAAGCAGCTGTGTGACAGAA 846  
QY 856 TTCAACGCGGTACACGCTCGCAACGCGATACCGCGACAGCGCGGAAACCTCAACCCAG 915  
Db 847 TTCCATAATGGACACTGGACACGCTCTGC---AGATCAGGCTGGAAACCTTTACATC 903  
QY 916 CAGGAAGTGCAGAGTTCTCTCGGCGCGGATGTTTCGGCCGTTTGAAGAGCGCGCGCG 975  
Db 904 GAGGAGGCGCTCTACTTTTGGGCTCTCGGATGTTTGGTCTGGAACAGTGGCGCACCA 963  
QY 976 ATTGACCTCGCGCGACGCGGCGACCCAGCGCTCGGCGCGACCCGAGAGGAAACAA 1035  
Db 964 ATCGATCTCACTCTGACGCTCGAGATCTTACCTTGGCAATGATCTTCAGCGGCANCAAC 1023  
QY 1036 AATTTTCGATTAAC-----TCAGACAGCTGACGGACGAGACCGCTGCGCCCTTC 1083  
Db 1024 GACTTCAACTACATACATCCAGGAGAAGACCTTACGACAGACGAAACCCGCTGCCCGTTC 1083  
QY 1084 GGTGCACACGTTGAGGAAGACGAAACCTCTCG-----ACAGGACCTGGGTGGACCGTCCGAC 1137  
Db 1084 ACTGCACACGTTCCAGACACCAACCGCGTGATTTAGAGACACAGGCTTTATCCCTGAT 1143  
QY 1138 ACCTTTCACGCTATGCGGTCAATGATCCGTCAGCGGCCAGAAACCTCTGATGACAGAACTT 1197  
Db 1144 CTGTTTCCATGTCATCCGTCGGGAAACGCTTATGGGCTTGAGGTTTACCGATGCCGAAATCC 1203  
QY 1198 GGTGCGGCGTGAAGTGGCAAGACCGCGGTCTTTCTTCGTCGAGTACCAAGTCCATTAAT 1257  
Db 1204 AACAGCACACTACTTTCGATCGATCTGCTGGCATTTGTGAGTACCAAGTCCGCTCAT 1263  
QY 1258 GGTAAATGGGTTTCAAGGTTTCCAGCAGATTAATCTGGGCAACAAATGCGAACTTCCCTTTCTCC 1317  
Db 1264 TCTAATGGATTCGATTTCCAGCAGCTCAATTTGGGCCAACCAACCGCAATTTCCCTTCAAC 1323  
QY 1318 AAAACCGATCAGCGCTGGAATTTGAGCTATCATCTCGGCAGACGACTCCACGACATCTCGGC 1377  
Db 1324 AAGTCTGAGCCACTCGGACTTGATCCGCTTATCGGACAGGACTCTGCCAGACATTCGCG 1383  
QY 1378 GGGCTCGACCCCTCAACCGAATGACAGCTTACAGTACCGCTGTTTGTGATCCGGAAG 1437  
Db 1384 ---TTGGACCGCGAAACGCGCTCGGATAGTCTCAGCATTCCTCAGATAATCATCTCTAAC 1440  
QY 1438 GCGGGGAATACTTTTCTTTCGCTCTATCTCTGCGCTC 1476  
Db 1441 GGAGCGAATATTTCTTCTCTCCATCTATCAGAGCCCTC 1479  
RESULT 4  
AF468656  
LOCUS  
DEFINITION Termitomyces albuminosus peroxidase TAP mRNA, complete cds.  
ACCESSION AF468656  
VERSION AF468656.1 GI:20386143  
KEYWORDS  
SOURCE Termitomyces albuminosus.  
ORGANISM Termitomyces albuminosus

Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
Agaricales; Tricholomataceae; Termitomyces.  
REFERENCE  
1 (bases 1 to 1735)  
Johjima, T., Okuma, M., and Kudo, T.  
TITLE  
JOURNAL  
Unpublished  
2 (bases 1 to 1735)  
Johjima, T., Okuma, M., and Kudo, T.  
REFERENCE  
AUTHORS  
Johjima, T., Okuma, M., and Kudo, T.  
TITLE  
JOURNAL  
Direct Submission  
Submitted (14-JAN-2002) JST Biorecycle project, Kasetsart Research  
and Development Institute, Bangkok 10900, Thailand  
Location/Qualifiers

## FEATURES

source  
1..1735  
/organism="Termitomyces albuminosus"  
/strain="ATCC42010"  
/db\_xref="ATCC:42010"  
/db\_xref="taxon:79931"  
34..1548  
/codon\_start=1  
/product="peroxidase TAP"  
/protein\_id="TAM21606.1"  
/db\_xref="GI:2038614"  
/translation="MOLKFLAATAAFVALSOPIAHYKARASSPLIVPPGQALP  
TAAQVSTASGDSIPRENTQADILVGMKQKKEKPFPHINAAEFKSLVTYAPNT  
TSVALTIGPISAQPAVNVAFSAQAGLGLTNDLGDPAFTGCPAPASDGLDDPN  
TWEPFKNTIDYFLISDSDITQYODDQAKGDMTYVYVLSGARPDQFCH  
EHRGLDSISNPLKGFGEPLDQFLDREIILVGRANDTYTRPAMLDSEFLAKK  
LKQJLPEHFKYLLNALQNSGNSLSTEEGALLSGSMFPMNSGAPIDLTPVDVPL  
GNDPRNNNFYIHPGEPDPAIDSRCPFAHIRKTNPLRLEQNLIPFPAIIRAGTP  
YGPVSVASSSNTTQIDRGIAFVYOSINSNFRQOVNMANNNKPFNKSISEPGL  
DPVIGODGARIRITGLNPLNVSLSIPDFIVSNGEYFAPSITSIVETLAA"

## CDS

BASE COUNT 437 a 483 c 371 g 444 t  
ORIGIN  
Query Match 21.8%; Score 325.6; DB 8; Length 1735;  
Best Local Similarity 56.2%; Pred. No. 3.2e-62;  
Matches 811; Conservative 0; Mismatches 589; Indels 42; Gaps 9;

85 CAGACCGGTACGACACCCCTCTCACTAATTTTCGGGACAAAGCCCGTGGCACTCTA 144  
Db 112 CGGGCAAGGTACTCTCCCTCATCGTCCCTTCCCTGGCCAGGGGGACATCAACATGA 171  
Qy 145 ACCGACATACGATCGAGAGCGGGGCCAACATACATTTGCCCCGAAACAATACAA 204  
Db 172 GCCCAGGTTCATC--AACCTTCGCGCGCATGACAGCAATCCCTTCGAAATATTGAG 228  
Qy 205 GCGGACATTTGGTGGCATGAAAGAACAGAGAAAGCGCTTCCTTTTCCAAATGCAT 264  
Db 229 GCAGATATCTGTGGGTATGAAGAGAAAGAAATTTGTCTTCTTCAATCAAC 288  
Qy 265 GACGCAACTCGTTCAAGAGCGGCTTGAAGACTTACGCTCGGACGCATACAGTGGCG 324  
Db 289 AACGCGCCACGTTCAAAAGCGTTCTCAAGACATATGCTCGGCTAATATTACCTCGTT 348  
Qy 325 GCGATTTGATTTAGATCTCTCAGAGCGGTTGCTTCGCAACCTCGGATTTTGG 384  
Db 349 GCTACCTTATAGTCCCATCTCCGCCAACCAAGAGATGATTAACCACTTGGGAGCTGCT 408  
Qy 385 AACACAGGCTCCAGCGCTTGAATTAACCAAGATCTGGGTATGCAAAATTCACAGAT 444  
Db 409 CAGGCTGGGTGGGTGCACTGGGTGATGATATCTCGCGCACCTGCTTACCTGGG 468  
Qy 445 GGTCAAGTTCGACAGCGCGCAAA---CTCGGGGACGACCTCAGCCAAATGGGTGGCGCT 501  
Db 469 GGGCAGTTCCGCTGATGCTCCGATGCTTGGGATGATATACCACTTGGGAGCTGCT 528  
Qy 502 TTTACTGTGTAACCACTCATGCTCTTTCTGATGGTACGACCAAGCACTTCTTG 561  
Db 529 TTTCAAAAGGCAACATATTAAGCGCGTTTCTGATCATATGACAAAGATTCAATTAAT 588  
Qy 562 GATCAGTTACGAGATATCTCTTGCAGCTTGGTTCCTCATCACTCAGGTGACGGCG 621  
Db 589 ACTCAGTACCAAGATATCTCCAGGCAAGCTAGGAGAGCGGTGATCTGCTATAGAT 648

Qy 622 CTCAGTGGGTCTGGCGGCTCCAGAGATCAGGCTGTGATGAAACATCTGGGTTCTCCAG 681  
Db 649 TTGAGTGGGCGACGACGACCTGGGACCAAGTTGTGTCAGAAATTTGGGCTATCGAGC 708  
Qy 682 GGCATCTGCAAGCCCTCAGTCAAGGCTGGGAGAACAGCGCTTCCCTGGACAGCGGTC 741  
Db 709 GGTATCTCAACCTCTCATTAAGATTTGGGAAAC---CTTGCTGGCCAGACCTTT 765  
Qy 742 GTCCACCTCGGAATTAATCTCATCTGACGCGGATGGGAGACAGCGGAC---GACCGTCG 798  
Db 766 ATTGATTCGGGTATCATTTATGATGCGCGGCGCAATGATACATCACTGCGCCCGG 825  
Qy 799 TGGGCTAGATGGAGTTTCATGAGCATTCGCGACCTTCAGACGAAGTCCCGCAATTC 858  
Db 826 TGGGCTCTGATGTTTATCTCTCGCTTCGCAAGCTCAAAACGCTTGTCCGAAATTT 885  
Qy 859 AACGGTACAGCTTGGCCAAAGCGATACCGCGCAACAGCGCGGAAACTTACCAAGCAG 918  
Db 886 CACAGGTATACGCTAGCAATATGCTCTGCA--AACCAATCAGGAAACCTTTCAACCGAA 942  
Qy 919 GAAAGTGCAGAGTTCTCTGGCGCGCATGTTCGCGCTTGAAGAGCGCGCGCATTT 978  
Db 943 GAAAGTGTCTCTTTTGGGATCAAGGATGTTTGTGCTGAGAACGCGGTGCACCAATTC 1002  
Qy 979 GACCTCGCGCCGACGCGGACGACGACCTCGCGCCGACCCGACAGAAACAAT 1038  
Db 1003 GATCTAACCCCGATGTGATGATCCGCTCTGGCAACGACCTTAAACGAAACAAT 1062  
Qy 1039 TTGCAATTA-----CTCAGACGCTGACGCGACGAGCGGCTGCCCTTCGCT 1086  
Db 1063 TTCAACTATATCATCTCCGAGAAAGATCTGCGACACAGACGCGGTTGCTTTCACG 1122  
Qy 1087 GCACAGTGTAGAAACGAAACCC-----TCGACAGACCTGGTGGACCGGTCGACAC 1140  
Db 1123 GCGCAGATTCGAGAACCAATCTCCCGGATCTGGAATCTGAAATCTATCCCGCAATTC 1182  
Qy 1141 TTCACGCTATCGGTCAGATCCGTCAGGCGCCACAGAAAGCTGTGATCAGAACTTGG 1200  
Db 1183 TTCACGCAATTCGTGACGAAACACTTAAGAACAGAGTATCTGATCCGAGTCCAGC 1242  
Qy 1201 TCGGGGTATCTGGCAGACGCGGCTTCTTTTGTGAGTACCAAGTCAATATGTT 1260  
Db 1243 TCTAATACCAACCAATATGATCGGCTGTGCAATTTGTAGTACCAATCAATATCA 1302  
Qy 1261 AATGAGTTACGTTCCAGCAGATTAATCGGCGCAACCAATCGAACTTCCCTTTCGAA 1320  
Db 1303 AACGTTTTCGTTCCAAACAGTCAATGGCTAATTAACAAGATTCCTCCCTTCAATAG 1362  
Qy 1321 CGATATAC---GCTGGAATTTGAGCTATCATGGCCAGACGA-----CTCAGCGCACT 1371  
Db 1363 AGTATCTCGAACCGGCGCTTGACCTGTGATGGAAGAGTAAAGGTGCAACCGCAAT 1422  
Qy 1372 GTGCGGCGCTCGACCCCTCAACAGAAATGAGAGCTTCAAGTACCGGCTTGTGATC 1431  
Db 1423 ACAACGCGTTGAATCCCTCAACGTTTGGAGAGTTTGAAGATTCGGAATTCATCGTC 1482  
Qy 1432 CCGAAGGCGGGAATATCTTTTCTTGCCCTTATCTCTGCGCTCACTGCAATATCGCT 1491  
Db 1483 TCCAAATGGCGGGAATATCTTCTGCTCTTATCATATTAATGAGAGCTTTGCC 1542  
Qy 1492 GC 1493  
Db 1543 GC 1544

RESULT 5  
AF429315 125020 bp DNA linear PRI 18-JAN-2002  
LOCUS AF429315  
DEFINITION Homo sapiens junctional protein 3 (JPH3) gene, partial cds.  
ACCESSION AF429315  
VERSION AF429315.1 GI:17646244  
KEYWORDS



repeat\_region complement(35581..35746)  
/rpt\_type=tandem  
/rpt\_unit=ctg  
complement(36507..36887)  
/gene="UjP3"  
/note="UjP3"  
complement(36507..36887)  
/gene="UjP3"  
/product="junctional protein 3"  
/note="component of the junctional complex between plasma membrane and endoplasmic reticulum"  
/codon\_start=1  
/protein\_id="AAL0941.1"  
/db\_xref="GI:17646245"  
/translation="MSSGGRFNPDDGSGYCGHEQKAGHGVCTPGPGGEYTGSMHGFVGLGTYTPSGNTYOGTWAQKTHGIGLESKRWYKGEHTHGFGRYGRVRECAHNGARVETWSGLDGYGTETYSYG"  
BASE COUNT 29056 a 32731 c 30696 g 28283 t 4254 others  
ORIGIN  
Query Match 3.7%; Score 54.6; DB 9; Length 125020;  
Best Local Similarity 11.7%; Pred. No. 0.098;  
Matches 99; Conservative 363; Mismatches 366; Indels 18; Gaps 3;  
QY 204 AGCGCATTGTTGGTGGATGAAGAAACAGAGAAACGCTTCCTTTTCCAGTCAA 263  
DB 17778 RGGTGYMKKKKGGKSGSSKMKKSSSTRRRGSAKSCSYMGMSKMSMAKSYMM 17719  
QY 264 TGACGCACTGTTCAAGAGCGGGGCTTAAGACCTAGCTGCTGAGCGCATCAGTCCGCG 323  
DB 17718 CTCYATRMSYMSYKCYKSCMSMSYSYSCCMKMSMSGCTCKMKYISGMSYSTSM 17659  
QY 324 GCGATTGATTGATTCAGATCTTCTCAGACGCGCTTCCTGTCACCTCGGTTTTC 383  
DB 17658 GYSSSTSCCKYKCSMSWCKTKTSKSRKRSYYWGGGRKAKKYYCAGRRRMSYMKC 17599  
QY 384 GAACACAGCGCTCCAGGGGCTTGGAATTAACGAGATCTGGGTATGCACATCCCA 443  
DB 17598 CAKMMSYCMMSYCMYTYYSKCTYKSTCKRGGYWGSKTCYSAGKSRMYCMMS 17539  
QY 444 TGGTCACTGCGAGACGCGCAACCTG-----GGACGACCTCAGCAATGGGTG 496  
DB 17538 SKSSWSMSMASSWCMGAGYRSKRSAGWAGWSSSKKSTGKRAKSTGYS 17479  
QY 497 CGCTTTTACTGTACCAACCATCATGTGCTTTGATTTGATGAGGACGACGAGACT 556  
DB 17478 TGRSMKKKGYKYSRBMKKKKTKYCMKTKYKTKTSKCMYWMYMSWGYKTKRCKMK 17419  
QY 557 TCTTGATCAGTTACGATGATATCTTGCACCTTGGTTCCTCAT-----C 606  
DB 17418 KCGTGRGMSKSKSGYKSRGSRKSYSTSKMSKMSWMMCKYMSYKRRSMRSGM 17359  
QY 607 ACTCAGCTCAGCGCGCTCAGTGGGTCTGGCGGTCCAGAGATCAGTGGTCTGAAC 666  
DB 17358 SSKMGYAGRCGCTSSSMWSTKRSKCYSTYKYGKRGKMGKRSKYSKMSKMK 17299  
QY 667 TTCGGTTCTGAGCGGATCTCGAGGCTTCAGTCAAG-GCTGGAGACGACCGCTTT 725  
DB 17298 KMSMSKCYSTYSKSGRKSSTKRSKAKSMAGSKCTYSSSYNNRRNNMG 17239  
QY 726 CCGTGACGAGGGGTGCTCCACCTGGAATTAATCTCAGTGGAGCGGATGGAGACAGG 785  
DB 17238 KTCGNTMYRSRMMNMAAGCTTCCCANITGGGGAAGAGGCGSASRASCYKGR 17179  
QY 786 CACCCGACCGTCTGGGCTCTAGATGAGATTTCAGCATTCGCGACTTCAGACAGA 845  
DB 17178 MSKSKCYRSGTTRRCKRSKRGSRKSMGMRGSGKTSYKSGRGCYCMGMRK 17119  
QY 846 GGTCCCGAATTCAACGGGTACCGCTGCGCAAGCGATACCCCGGAACCGCGGGAGA 905

DB 17118 CTRCNSRRKMKYKRYRKRKMTCKMKYSMTYMATCRKCMCKKSCCCGCKMSMS 17059  
QY 906 CTTCAACCCAGCAGAGATGTCAGAGTTCTCTGGCGCGCGCATTCGCGCTTGAAGAG 965  
DB 17058 YSYSGYKSYMKMSYMMRSYSKSTSKRMSKRGMTGGRYKGGRSYKSGGSGYKGG 16999  
QY 965 CCGCGCGCGGATTCCTCGCGCGCGAGCAGACCGCGCTGGGGCGGACCGCA 1025  
DB 16998 SMGKRGKSTRESSAGKSSYKMSKCAKYSMSYSKCYRMMWCMYTYCMGCKMSSC 16939  
QY 1026 GAGGAA 1031  
DB 16938 WWSMR 16933  
RESULT 7  
AE001862  
LOCUS 213732 bp DNA linear BCT 22-NOV-1999  
DEFINITION Deinococcus radiodurans RI section 1 of 2 of the complete chromosome 2.  
ACCESSION AE001862 AE001825  
VERSION AE001862.1 GI:5460468  
KEYWORDS  
SOURCE  
ORGANISM  
Deinococcus radiodurans.  
Deinococcus radiodurans.  
Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;  
Deinococcaceae; Deinococcus.  
1 (bases 1 to 213732)  
White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,  
Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,  
Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,  
Vamathevan, J.J., Lam, P., McDonald, L., Uetrecht, T., Zalewski, C.,  
Makarov, K.S., Aravind, L., Daly, M.J., Fraser, C.M., et al.  
Genome sequence of the radioresistant bacterium Deinococcus  
radiodurans RI  
Science 286 (5444), 1571-1577 (1999)  
2 (bases 1 to 213732)  
White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,  
Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,  
Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,  
Vamathevan, J.J., Lam, P., McDonald, L., Uetrecht, T., Zalewski, C.,  
Makarov, K.S., Aravind, L., Daly, M.J., Minton, K.W.,  
Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S.,  
Smith, H.O., Venter, J.C. and Fraser, C.M.  
Direct Submission  
Submitted (02-NOV-1999) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
LOCATION/Qualifiers  
1. 213732  
/organism="Deinococcus radiodurans"  
/strain="R1"  
/db\_xref="taxon:1299"  
/chromosome="2"  
653..1435  
/gene="DRA0001"  
653..1435  
/gene="DRA0001"  
/note="similar to GB:D26185.SP:D37522.PID:467381  
PID:580906.GB:AL009126 percent identity: 54.96; identified  
by sequence similarity; putative"  
/codon\_start=1  
/product="chromosome partitioning ATPase, putative, ParA  
family"  
/protein\_id="AAFI2301.1"  
/db\_xref="GI:6460595"  
/translation="MWSAVKTLTFVNHAGAKTSLTINVGYELARGSLRVLLDLP  
OANTLGLIGSGTREVTVPVVDGQPLPSKFAFGIDVPAHVSILAVAGOMGGRV  
GAGRLRALAEVSGDYDALYDFSPSLGLAIFLAALADOMIVPVPRGGLALPG  
LQGLATREVRBDLYALYVPIFYARRRRHDEVLADLKAHLSLAFVPRGRANVH



gene	DSTAOGAPVSEYAPGTPVHADVQRLTADIAAAIGVAYPGENA" 1432. .2313 /gene="DRA0002" CDS 1432. .2313 /gene="DRA0002" /notes="similar to GB:L34077 GB:X85964 PID:407375 PID:757761 PID:I163135 percent identity: 57.34; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="chromosome partitioning protein, ParB family" /protein_id="AAF12300.1" /db_xref="GI:6460594" /translation="MTKRRPERRRLLGLGETPVDLSQANDIRALPVNELKVGSTOP RRSFDERISELAESIRAHGVLOPLLVRSDVQVEIVAGERRRWAQAOLAGLAVPVV RQISNQARAAALINENRDNLNVDVQKLELIATLGLEREAKRMLQLLRVVP GDSHEQIDQVFRSMGTWTFANKRLRLNMPQVLEALRAGLPLTIGSVVAGPPER QALLLAONGASRSQLQALQTSQTSAVTPEHFAKVLSSKRFLSLGLDPTPREALDR WLAMPERVQAIDQES" complement (2418. .3881) /gene="DRA0003" CDS complement (2418. .3881) /gene="DRA0003" /notes="similar to GB:M88334 SP:P25526 PID:147901 GB:U00096 PID:1789015 percent identity: 69.60; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="succinate-semialdehyde dehydrogenase" /protein_id="AAF12294.1" /db_xref="GI:6460588" /translation="MTTDLRTTYTSVTRSQAYFDGEWRNAPRNFVRVHPGNEVIGE VACQTTDRAQIDAARLREWRQNPYERKILRRWHDLMPEHKEELAQLMTLEM KPISETRGEVHYAAEFIEWCASEAGRIAGERINLRFPHKRLGISERVGVYVATPW FPAGMTKRAAPALAGCAVILKPAELSPMTALYLTELKAGGPANTFOVLPNTDAS ALTQPMNDSRVKLTFTGSTEVLVYQAAAGTIKRVSLGELGGHAPLPVDDADLER AASEVASKRNSGQTCVTNRYVQGVASEFIRLLTEKTAALQLGSDPEDEATQVP VYQAGLDQVQDALTKQAQNTIGQVSSGLFFQPTVLVDVAPDSLLIRETQGP VAPVPTDEEGLRLANDSEYGLAAAYATRLDGRAPRIAGLEYGIVGINDDGUPSSA APHVFGMKNSGVREGHWGLEBYLETKFVSLGS" 4024. .4668 /gene="DRA0004" CDS 4024. .4668 /gene="DRA0004" /notes="similar to GP:2924378 percent identity: 55.29; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAF12299.1" /db_xref="GI:6460593" /translation="MNSLSLAAVLMLMAGSTRMGOPKALLPLGGMPLCRHAAQTIVE AGYGEWAVLPCEGTGDVRAALQGLPURWNTNPOPAQGLSSFQRAAGEALMAESAAP AVVFTLADNPLVRDRTHQALGVFGAEQPAALITRYGEVAAPPTLLRRDLILPLHEFP PSDHGPTLLREUGRELKIDRAPAAELLDITDIPENAAQQQLFAEVTFGAQDGP" complement (4719. .5897) /gene="DRA0005" CDS complement (4719. .5897) /gene="DRA0005" /notes="similar to GB:L33464 SP:P47734 PID:496118 percent identity: 74.23; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="alcohol dehydrogenase, zinc-containing" /protein_id="AAF12293.1" /db_xref="GI:6460587" /translation="MKAIWQGVNKGVTVPDPTLLLPDIAIKVITSTACGSDLHL LDGYIPSMETGDLIGHEPFWGEVVEVSGVRKLVKGDVWVPFNIAACVCPCKRGRFFS ACDNSPNHRMAALYCATSGGALPGYSHMYGGYAGGAQOAVRVVPFADVPFKIESGL KDSQVLTFIDFTFQAEEQCNIVPGRDVVAVFGVPGQVFAARSQAQMGLAAHVIYI DRVPERLMAAGCQTINYEKEDVLLSREATGGRGPDHVDVAGMEAHGHPGSTM DKVQONKLTDFDRITRALWAILSCAKGGTVSNPGVYGLVDKMPIGAFAKGLIFRMG
gene	complement (5894. .6700) /gene="DRA0006" CDS complement (5894. .6700) /gene="DRA0006" /notes="similar to PID:1653488 percent identity: 48.46; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAF12292.1" /db_xref="GI:6460586" /translation="MDRRSSPOLRTWNTWNEQOPAPHAGATGNSSPCATMQPLERSL TGAVTAAALAAAGLRSGLEKVLTLGAGAGIAVAARQNPALATKIEQNDGETVV EDAVTIGKPAEQLYAVWRDLPGLLPMTHLSREVLDKRSRMTVEAPAPLGTVSWEA ELTADPGRRIARWLPGARINSEGEVLFPAPGARCTEVVVRLTYRPPGGSAGAVIA RMFNQPPSQLDRLMRFKREGLCHAPTAQOQSSGRAALTTPNKRGLV" 6839. .7744 /gene="DRA0007" CDS 6839. .7744 /gene="DRA0007" /notes="similar to GB:AL009126 percent identity: 62.98; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="ABC transporter, ATP-binding protein, putative" /protein_id="AAF12291.1" /db_xref="GI:6460585" /translation="MITTFEQVSKTYGHVTALSDFNLTSLTGBELTALLGPNAGAKSTA IGLLGLSAPGQVRVIGFGRNDNRKARIGAMPQESALPAGLVREAVTLFSTFP APIGDEALADLADLPGVARRAAQSGQKRLAFVAVDGPDLLEDIPPTGMDAQ SRAAFVEVLAARGLTILTHYLEBEATADRVVMNGGRILADDITPQGLRSGVG GARVSFVSLVQAELERLPGVSAYQVDAAGRADLRTSVPEALLAALIGSGTTSFQLE RRATLEAVLQLTGPDMTAVTRSA" 7861. .8652 /gene="DRA0008" CDS 7861. .8652 /gene="DRA0008" /notes="similar to GB:AL009126 percent identity: 52.46; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAF12290.1" /db_xref="GI:6460584" /translation="MTTLFNPVPAKRTSALPVLAQLTLGELRRLRLRSPMTVGAIGFPV MFFALFGLPAVEYGAATDPHGVPILOFAAYSLLSLALFSGAAVATERSGGWRLRL RSSPLFVPLYFVSKTALALAFGAVSLALLYAFAPFAGGVTLPGLALLAGKLLGMI PLVALGICIGFLASPAAQILANILSVVMSPASGLFVPLDQLPGFVQOIAPLLPAVHV SQIATNTVSGQTASEPAHWLAAPTLVFGTLAVMGLKROESREG" 8809. .9819 /gene="DRA0009" CDS 8809. .9819 /gene="DRA0009" /notes="similar to GB:AL009126 percent identity: 55.80; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAF12289.1" /db_xref="GI:6460583" /translation="MLFNGTLVAFVAVVYVTFGKRDPPDDRRKAYVWLSALLMHALGDM LYGPSALSLFVYGGSLIGWRRQWVALGGAANALLMWGRLLTLPERSQGDIFILPLFI LASAYANHATYGLRARRLAQVLEKEKLAADAEERIERIARDLHDLIGHTLSVILKLS ELAGKLAKEHPQRAAEIREVERISREALSVRAAVQVQSGSLAELAKVGLDRA GIRLVITDTPLELPALESTLVLRVREAVTNVVRHSGASELRILTQKANGFOLTQD DGRGGAPEGCTGLNSMRERLRAMGCTFERDGHGTQTRLOAFLVPLSAGTAGLKAPELLEK TS" 9816. .10451 /gene="DRA0010" CDS 9816. .10451 /gene="DRA0010"



Query Match 3.7%; Score 54.6; DB 1; Length 213732;  
 Best Local Similarity 50.1%; Pred. No. 0.099;  
 Matches 197; Conservative 0; Mismatches 184; Indels 12; Gaps 2;

QY 925 GCAGAGTCTCTGGGCGCGCATGTTGGCCGTTGGAAGAGCGCGCGCGCATTTACCTC 984  
 Db 144981 GCCGAGCTCTTGCGCCCAAACTGGTGGAGCGCTGGCGCGCGCGCGCGCATTTG 145040

QY 985 GCGCGGAGCGCGGAGCGAGCCAGCGCTGGCGCGCGCGCGGAGGAAACATTTTCAT 1044  
 Db 145041 GCTCCGAGGAG 145100

QY 1045 TACTCAGACAGCTGAGCGAGAGAGAGCGCGCTGCGCTTGGTGACACAGTGAAGAGAG 1104  
 Db 145101 TACAAAAG 145157

QY 1105 AACCTTCAG 1164  
 Db 145158 AACCTTCAG 145217

QY 1165 CCGTACGCGCGCGAG 1224  
 Db 145218 CCGCGACGCGCGCTATAGCGCGCGCTATAGAGAGAGAGAGAGAGAGAGAGAGAG 145277

QY 1225 GGTCTCTTTTGTGAGATACAGAGATTTTGTGATAGGTT-----CAGGTTG 1275  
 Db 145278 GTTGAAGCTGGCTCTTACTTATCTTCTTCAAGCGCGAGAGAGAGAGAGAGAGAG 145337

QY 1276 CAGCAGATTAACTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1308  
 Db 145338 CTGCAAGAGAGATGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 145370

RESULT 9  
 AY034474/c 2040 bp DNA linear BCT 10-JUN-2001

LOCUS Methylobacterium extorquens dioxxygenase gene, complete cds.  
 DEFINITION AY034474  
 ACCESSION AY034474.1 GI:14334396

KEYWORDS Methylobacterium extorquens.  
 Methylobacterium extorquens.  
 Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 Methylobacterium group; Methylobacterium.

REFERENCE 1 (bases 1 to 2040)  
 Toyama,H., Matsushita,K., Adachi,O. and Lidstrom,M.E.  
 Cloning and analysis of the dioxxygenase gene in downstream of pqfG  
 Unpublished  
 2 (bases 1 to 2040)  
 Toyama,H., Matsushita,K., Adachi,O. and Lidstrom,M.E.  
 Direct Submission  
 Submitted (11-MAY-2001) Biological Chemistry, Yamaguchi University,  
 1677-1 Yoshida, Yamaguchi 753-8515, Japan

FEATURES  
 Location/Qualifiers  
 1..2040  
 /organism="Methylobacterium extorquens"  
 /db\_xref="taxon:409"  
 1..933  
 /note="orf88"  
 /codon\_start=1  
 /transl\_table=1  
 /product="dioxxygenase"  
 /protein\_id="AAK59392.1"  
 /db\_xref="GI:14334397"  
 /translation="MQLTGLHTATTAQAAADLAFTYVULGLRLVKTUNODVSNV  
 HLTADGASPGDITTFDMPAPAEKGTDSIRTLIRGGASGDMKREHLAQGS  
 HHPAIERGRITLDFEDGEGQRLSLVHDGQVPAHPWAGSPVPEHQIRGLGRLIV  
 SQPRTAEVILTVQVIGFRRARTFELPEGEVTEPVGGAALVQLKDTGA.PAQGAG  
 AVHVAPEIPDADYADAWADRLKQRVSSGPVDYRFSLYFERNGLIFBIATDGRG  
 FATDETMSLQERLALPPLEPRRAIEIGLKPL"

BASE COUNT 319 a 704 c 674 g 343 t

Query Match 3.5%; Score 52; DB 1; Length 2040;  
 Best Local Similarity 59.5%; Pred. No. 0.36;  
 Matches 88; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 1208 TCACTGCGAAGACCGCGGTCTTTTGTGTGAGTACAGTTCATTATGGTATGAGT 1267  
 Db 1273 TCCCAAGAGAGCGCTTCGAGCTGTTTCATGTGCTTCAATTCAAAATTTGCCACAGT 1214

QY 1268 TCAGGTTCCAGAGATTAACTGGGGAACAGTTCCTTCTCCTCAACAGGATCA 1327  
 Db 1213 TGGCGTTCCAGGAGCAATTGGGGAACAGTTCCTTCTCCTCAACAGGATCA 1154

QY 1328 CCGCTGGAATTGAGCTTATCATTCGCGCA 1355  
 Db 1153 CCCATGAGTGCATCCGGTATCGGCCA 1126

RESULT 9  
 SCF42/c 34713 bp DNA linear BCT 12-MAY-2002

LOCUS Streptomyces coelicolor cosmid F42.  
 DEFINITION AL137165 AL645882  
 ACCESSION AL137165.2 GI:20520910

KEYWORDS ABC transporter protein, ATP-binding component, AraC-family  
 regulator, branched-chain amino acid transport protein;  
 branched-chain amino acid transport ATP-binding protein;  
 DNA-binding protein; glycosyl hydrolase; glycosyl transferase;  
 hydrolase; integral membrane protein; lipA; lipase; lipoprotein;  
 lipR; membrane protein; oxidoreductase; transcriptional activator.  
 Streptomyces coelicolor A3(2).  
 Streptomyces coelicolor A3(2).  
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1 (bases 1 to 34713)  
 Regenbach,M., Krieser,H.M., Denapate,D., Eichner,A., Cullum,J.,  
 Kinashi,H. and Hopwood,A.  
 A set of ordered cosmids and a detailed genetic and physical map  
 for the 8 Mb Streptomyces coelicolor A3(2) chromosome  
 Mol. Microbiol. 21 (1), 77-96 (1996)  
 97000351

JOURNAL PUBLISHED  
 MEDLINE 884336

REFERENCE 2 (bases 1 to 34713)  
 Oliver,K. and Harris,D.  
 Unpublished  
 3 (bases 1 to 34713)  
 Thomson,N.R., Parkhill,J., Barrell,B.G. and Randleman,M.A.  
 Direct Submission  
 Submitted (24-JAN-2000) Streptomyces coelicolor sequencing project,  
 Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
 CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.  
 David A. Hopwood, [3] John Innes Centre, Norwich Research Park,  
 Colney, Norwich, Norfolk NR4 7UH, UK  
 On May 9, 2002 this sequence version replaced gi:6752308.

COMMENT  
 Note:  
 Streptomyces coelicolor sequencing at The Sanger Centre is funded  
 by the BBSRC and Beowulf Genomics  
 Details of S. coelicolor sequencing at the Sanger Centre are  
 available on the World Wide Web.  
 (URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>) CDS are  
 numbered using the following system eg SC787.01c. SC (S.  
 coelicolor), 787 (cosmid name), .01 (first CDS), c (complementary  
 strand).  
 The more significant matches with motifs in the PROSITE database  
 are also included but some of these may be fortuitous. The length  
 in codons is given for each CDS.  
 Usually the highest scoring match found by fasta -o is given for  
 CDS which show significant similarity to other CDS in the database.  
 The position of possible ribosome binding site sequences are given  
 where these have been used to deduce the initiation codon. Gene  
 prediction is based on positional base preference in codons using a  
 specially developed Hidden Markov Model (Krogh et al., Nucleic  
 Acids Research, 22(12):4768-4778(1994)) and the FramePlot program  
 of Bibb et al., Gene 30:157-66(1984)) as implemented at

<http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid F42 lies alongside and overlaps with cosmid F15 on the AseI-F genomic restriction fragment.

FEATURES

source	Location/Qualifiers
	1. .34713
	/organism="Streptomyces coelicolor A3(2)"
	/strain="A3(2)"
	/db_xref="taxon:100226"
	/clone="cosmid F42"
misc_feature	2. .892
	/note="pfam match to entry PF01315 Ald_Xan_dh_C, Aldehyde oxidase and xanthine dehydrogenase, C terminus, score -294.50, E-value 0.00012"
gene	/gene="SCF42.02c."
	/note="SCO0692"
CDS	complement(1187..2644)
	/gene="SCF42.02c."
	/note="SCF42.02c, possible membrane protein, len: 485 aa. Rich in the amino acid Ala. Contains possible membrane spanning hydrophobic domains."
	/codon_start=1
	/transl_table=11
	/product="putative membrane protein."
	/protein_id="CAB69664.1"
	/db_xref="GI:6752310"
	/db_xref="SPTRMBL:Q9L2L1"
	/translation="MTGSSSRSEYALISALDLTGTTTTTGTARPPGTEAEARAS SARPSEQSDIWDVDFDADLPEPEPDPDLPGLPAPDGGPGRALDADGAPAPAG ADADAQHGDDPDLVRAAVADRLDEVVDLITTLERSQYQAQTDALRAVGNRS VEDVTLGLTLTPPQSDSADAEIARAAECRSVEDVTRMLALHRTPLPFGCQAAV RAAATGPVELVELIGRLAESGHRPERDARLRADRDAAEDDGDGGRAPRSAGRG GTSAGADRGRERRPRFRGRDRDETAAARDRARDRARDRARDRARDRARDKARA VRDARRAARGAPAWLTVAVLACGVAVPPLHRGDASAPAYGVALGISALCLVLA LLLTVRPVPLMAAAVVGPAALAAKLYGSATPSPARTVFTDTLAPVWAVAVAA SLVALALCVRASQDSARRRPARPVVSSRTAD"
RBS	2816..2820
gene	2822..3169
	/gene="SCF42.03"
	/note="SCO0693"
CDS	2822..3169
	/gene="SCF42.03"
	/note="SCF42.03, possible membrane protein, len: 115 aa. Contains a possible N-terminal signal sequence and a possible membrane spanning hydrophobic domain. Rich in Ala and Leu."
	/codon_start=1
	/transl_table=11
	/product="putative membrane protein."
	/protein_id="CAB69665.1"
	/db_xref="GI:6752311"
	/db_xref="SPTRMBL:Q9L2L0"
	/translation="MKRAPEEGAVDTTVIVVAVULLAALVLAALAAVAVLVLRTRDRL RRRTGPGDYAA"
RBS	3280..3283
gene	3289..3480
	/gene="SCF42.04"
	/note="SCO0694"
CDS	3289..3480
	/gene="SCF42.04"
	/note="SCF42.04, unknown, len: 62 aa."
	/codon_start=1

	/transl_table=11
	/product="hypothetical protein SCF42.04."
	/protein_id="CAB69666.1"
	/db_xref="GI:6752312"
	/db_xref="SPTRMBL:Q9L2K9"
	/translation="MQPFALNYPAVELEATTFFVYDSGLQLNLLDGRVAACHAL LRLGTTTISAGSKTHFDD"
gene	3497..4483
	/gene="SCF42.05"
	/note="SCO0695"
CDS	3497..4483
	/gene="SCF42.05"
	/note="SCF42.05, unknown, len: 328 aa: similar to TR:Q9KZTD (EMBL:ALJ53832) hypothetical protein SCF6.08 from Streptomyces coelicolor (326 aa) fasta scores; opt: 371, Z-score: 407.1, E(): 4.8e-15, 35.629% identity (39.145% ungapped) in 334 aa overlap."
	/codon_start=1
	/transl_table=11
	/product="conserved hypothetical protein SCF42.05."
	/protein_id="CAB69667.1"
	/db_xref="GI:6752313"
	/db_xref="SPTRMBL:Q9L2K8"
	/translation="MTVLILTSEEDVTADVVVHLNAGVVPVRLDPADLTDSVALSG EFAHGSFRGLSSGRLVSIIGLRSVWVRPFGAATRAAPESAWLTEAQALYGLMR GSGARWNPDAARHARYKPKQLARQRCGLPVPATLITTPRAAREFAERYPDLVVK PVSGAHQDPPLAVPTSRVPEADPFSVAHGPRTLQRRVAKRADIRLTAVGELLAAAR KTAASLDDEVDVRFAGSGBEMRPAEVPVRVAGSVRAYLRAAGLAVGALDPAEDGDG TWHFSCNOSGOGFVEVDTSQPTARTIAEWLARPGADAVEGPGPDAAVG"
gene	complement(4823..5317)
	/gene="SCF42.06c"
	/note="SCO0696"
CDS	complement(4823..5317)
	/gene="SCF42.06c"
	/note="SCF42.06c, possible integral membrane protein, len: 164 aa. Contains multiple possible membrane spanning hydrophobic domains."
	/codon_start=1
	/transl_table=11
	/product="putative integral membrane protein."
	/protein_id="CAB69668.1"
	/db_xref="GI:6752314"
	/db_xref="SPTRMBL:Q9L2K7"
	/translation="MTKILLSLHLVLAALVAVGPVTVAASMPFAARRVPVAATVGAD GAAAGDAGGVTVRLHRRICRVYAGLGIAPVLGFATAAAGVLSDAWLVASITLTAV AAGLLAAFPVPRQELSELDELADGRPVERARTARLMTGVFNLLMATVTVLMVRFGS TTGA"
gene	5572..6459
	/gene="SCF42.07"
	/note="SCO0697"
CDS	pseudo
	5572..6459
	/gene="SCF42.07"
	/note="SCF42.07, possible AraC-family regulator, pseudogene, len: 309 aa. Similar to several other putative AraC-family transcriptional regulators from Streptomyces coelicolor e.g. TR:CAB61299 (EMBL:AL132991) SCF55.29 (311 aa) and TR:Q9XA73 (EMBL:AL096822) SCGD3.05 (334 aa). Contains a frameshift after codon 188, following a possible deletion event."
	/pseudo
	/codon_start=1
	/transl_table=11
	/product="putative AraC-family regulator (pseudogene)"
gene	complement(6487..6852)
	/gene="SCF42.08c"
	/note="SCO0698"
CDS	complement(6487..6852)
	/gene="SCF42.08c"
	/note="SCF42.08c, unknown, len: 121 aa."
	/codon_start=1
	/transl_table=11
	/product="hypothetical protein SCF42.08c."

Query Match 3.5%: Score 51.8; DB 1; Length 34713;  
Best Local Similarity 48.5%; Pred. No. 0.41;  
Matches 143; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

870 GCTGCGCAACGCGATATCCCGCAACAGCGCGGGAACCTTCACCCAGAGAGATGACAGA 929  
19596 GCTGCGCGCGCGCTCTCGACGCGACCGGTCGATCGCCGACCTCCCGAGCGCGCG 19537  
QY 930 GTTCTCGCGCGCGCGCATATTCGCGCTTGGAGAGAGGCGCGCCGATTCGCTGCGCG 989  
Db 19536 GCGCCCTCGCGCTCCCGAGACGCGCGCTGATCTGTGTCGCGCGCGCGCGACCC 19477  
QY 990 GACGCGGACGACGACCGCGCTCGCGCGCGACCGCGAGAGAGAACATTTTCGATTATC 1049  
Db 19476 GCG 19417  
QY 1050 AARACAGCTGACGAGAGAGAGAGCGCGCTTGTGTGTGACAGTGAAGAAGACGACCC 1109  
Db 19416 CGGCGGAGAGACGACGACGATGCTGTGCTGTGCGACGACGAGCGGCTCCGGAAC 19357  
QY 1110 TCGACAGACGCTGCGTGACCGCTGACACCTTCACGCTATGCGGTCCGATATC 1164  
Db 19356 GCACGCGCGCGAGCGCGCGAGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19302

RESULT 10  
DROACT88F  
LOCUS DROACT88F 2150 bp DNA linear INV 26-APR-1993  
DEFINITION D.melanogaster actin gene, complete cds, locus 88F.  
ACCESSION M18830 J01065  
VERSION M18830.1 GI:156772  
KEYWORDS actin.  
SOURCE Drosophila melanogaster (Canton S) DNA.  
ORGANISM Drosophila melanogaster.  
Eukaryota; Metazoa; Arthropoda; Insecta; Pezomyzeta;  
Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 2150)  
Sanchez, F., Tobin, S.L., Rdest, U., Zulauf, E. and McCarthy, B.J.  
Two Drosophila actin genes in detail. Gene structure, protein  
structure and transcription during development  
J. Mol. Biol. 163 (4), 533-551 (1983)  
6405041

COMMENT (1) also shows the complete sequence of the actin gene at  
cytological locus 79B and found that the 79B and 88F actin genes  
exhibit a combination of co-ordinate and differential patterns of  
messenger accumulation. They appear to follow parallel patterns of  
transcription through embryogenesis to the prepupal stage but  
subsequently differ greatly with regard to the proportion of  
message from each gene in pupae. Of the Drosophila actin introns  
studied so far, the intervening sequences of the 79B and 88F genes  
are the only ones inserted at the same position.

FEATURES  
source  
1..2150  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
prim\_transcript 496..21874  
/note="actin mRNA"  
CDS join(684..1608,1669..1874)  
/note="actin"  
/codon\_start=1  
/protein\_id="AA28321.1"  
/db\_xref="GI:156773"  
/translation="MCDPDAGLVINGSGMCKAGADADAPRAVPSIVGRPHGV  
MYMGQKDSYDGAQSKRGITLKPRIHRCITTKNDMEKIMHTTFNLRVADBEH  
FVLITAPLNKANKREKQITFETNSPLMVAIQAVSLASRTGIVLSDGDEV  
SHVIVIEGALPFAVILRLDLAGRDLPVIMKILTERGFTTTERIVDITKXIC  
EYVADFEDEVATAASTLSKSYELPDGQVITIGNERPCFPALPQSPFLQMECGIH  
ETVNSIMKCVADIRKDLVANSVLSGTTMYGIDRMQKEITTLAPSTIKIILIAP  
ERKTSVMIGLILASTFTQGMWISQKQVDESQPSIVHRKCP"

exon

BASE COUNT 544 a 562 c 517 g 527 t  
ORIGIN 77 bp upstream of HindI site, locus 88F.

Query Match 3.1%: Score 46.2; DB 3; Length 2150;  
Best Local Similarity 48.3%; Pred. No. 7.2;  
Matches 129; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

461 CCGCAACCTCGGGAGACCTTCAGCAATGAGGCGCCTTTACTGTATACCACTATCC 520  
Db 1243 CCGATTACTTATGAAATCTCTAGAGAGCGCGCTATACCTTCACACACCGCGACG 1302  
QY 521 ATGGTCTCTTCTGATTGTGAGCAGCAGACACTTTTGATCATGTTACGATGATA 580  
Db 1303 GTGAGATCGTGGCGACACCAAGAGAACTGTACTACGTGCTGTGAATTGAGAGCG 1362  
QY 581 TCTCTTGACCTTGTGTTCTCCATCACTCAGGAGCGGCGCTCACTGAGCGCTC 640  
Db 1363 AGATGGCAGCGCTGCGCGCTCCACCTCGCTGAGAGAGTGTACGAGCTGTGAGCGCC 1422  
QY 641 CAGAGATGAGCTGTGCTATACACTTCGAGTTCTCGACGACATCTCGACGCTCAG 700  
Db 1423 AGGTATACCACTTGGCAACGAGCGCTTCGCTGCCGAGGCGCTGTTCAGCCCTCGT 1482  
QY 701 TCACAGCTGTGGAGACGACCGCTTTCG 727  
Db 1483 TCTCGGATGGAGTGTGCGGCACTCC 1509

RESULT 11  
SCB66/c  
LOCUS SCB66/c 28527 bp DNA linear BCT 12-MAY-2002  
DEFINITION Streptomyces coelicolor cosmid E66.  
ACCESSION AL359989 AL645882  
VERSION AL359989.1 GI:8894754  
KEYWORDS ABC transporter ATP-binding subunit; acetyltransferase;  
dehydratase; dimethyladenosine transferase; fatty acid desaturase;  
galE, UDP-glucose 4-epimerase; galK, galactokinase; galT,  
galactose-1-phosphate uridylyltransferase; isopenicillin  
monophosphate kinase; MavR-family; membrane; ppc,  
phosphoenolpyruvate carboxylase; prs, ribose-phosphate  
pyrophosphokinase; pth, peptidyl-L-lysine hydrolase; ribosomal L25p  
family protein; secreted; sodium:solute symporter; Teer-family  
transcriptional regulator; trans-aconitate methyltransferase;  
two-component response regulator; UDP-N-acetylglucosamine  
pyrophosphorylase.  
Streptomyces coelicolor A3(2).  
Streptomyces coelicolor A3(2).  
Bacteria; Firmicutes; Actinobacteria; Streptomycetaceae; Streptomyces.  
1 (bases 1 to 28527)  
Redenbach, M., Kleser, H.M., Denapalte, D., Eichner, A., Cullum, J.,  
Kinashi, H. and Hopwood, D.A.  
A set of ordered cosmids and a detailed genetic and physical map  
for the 8 Mb Streptomyces coelicolor A3(2) chromosome  
Mol. Microbiol. 21 (1), 77-96 (1996)  
97000351  
8843436

REFERENCE  
2 (bases 1 to 28527)  
Seeger, K.J. and Harris, D.  
Unpublished  
JOURNAL  
3 (bases 1 to 28527)  
Bentley, S.D., Parkhill, J., Barrell, B.G. and Rajandream, M.A.  
Direct Submission  
JOURNAL  
Submitted (28-JUN-2000) Streptomyces coelicolor sequencing project,  
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.

David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

Notes:  
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics  
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.  
(URL: [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/)) CDS are numbered using the following system eg SC7B7.01c, SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.  
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or attc) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
IMPORTANT: this sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid B66 Lies between and overlaps cosmids E41 and E87 on the AseI-E genomic restriction fragment.

FEATURES

source	1..28527 /organism="Streptomyces coelicolor A3(2)" /strain="A3(2)" /db_xref="taxon:100226" /clone="cosmid B66" 1..107
misc_feature	/note="nominal overlap with Streptomyces coelicolor cosmid SCB66"
misc_feature	405...458 /note="Nominal overlap with Streptomyces coelicolor cosmid E41" /note="Pfam match to entry PF00132 hexapep, Bacterial transferase hexapeptide (four repeats), score 6.20, E-value 1.1e-02"
misc_feature	459...512 /note="Pfam match to entry PF00132 hexapep, Bacterial transferase hexapeptide (four repeats), score 8.40, E-value 55"
misc_feature	558...611 /note="Pfam match to entry PF00132 hexapep, Bacterial transferase hexapeptide (four repeats), score 7.90, E-value 65"
misc_feature	663...716 /note="Pfam match to entry PF00132 hexapep, Bacterial transferase hexapeptide (four repeats), score 3.80, E-value 2.3e-02"
misc_feature	843...896 /note="Pfam match to entry PF00132 hexapep, Bacterial transferase hexapeptide (four repeats), score 6.80, E-value 93"
RBS	1153...1157
gene	1166...2140 /gene="SCB66.02" /note="SCO3123; prsA2" 1166...2140 /gene="SCB66.02" /note="SCB66.02, prsA2, ribose-phosphate pyrophosphokinase (EC 2.7.6.1), len: 324 aa; strongly similar to many eg.
CDS	

SW:P14193 (KPRS\_BACSU) ribose-phosphate pyrophosphokinase from Bacillus subtilis (317 aa) fasta scores; opt: 1000, z-score: 1178.1, E(): 0, 46.8% identity in 314 aa overlap. Contains Pfam match to entry PF00156 Priboyltran, Phosphoribosyl transferase domain and Prosite match to PS00103 Purine/pyrimidine phosphoribosyl transferases signature."

/codon\_start=1  
/transl\_table=11  
/product="ribose-phosphate pyrophosphokinase"  
/protein\_id="CAB95916.1"  
/db\_xref="GI:8994756"  
/db\_xref="SPTREMBL:Q9K3U0"  
/translation="WTGKTKTGEKMMFPFSORAHPELAEEVAQQLGVGVVPTKAFDPA NGSIYRYSESARGADCTLIQSHTAPKPNKMWQELIMIDALKKASAKSITVVPFGY DYQKKHREPREISARLIADLMKTAGADRIILAVDLHTDQIQGFDDGVGDLFALPDLA DYQKADRSKLTIVVPDAGRVADRCWCDRLGAPLAIVHKRDKDYANQVTVHEVVG DVGRICTLVDDMIDTGTTCAAADALFAHAEDVIIVTATHGLSGPAADRLKNSKVS EFTVFTLPSASELELDKITVLSIAPTITARAREVFEDGSGVTSLFDEQ"  
1592...2008  
/gene="SCE66.02"  
/note="Pfam match to entry PF00156 Priboyltran, Phosphoribosyl transferase domain, score 100.60, E-value 3e-26"  
1838...1876  
/gene="SCE66.02"  
/note="PS00103 Purine/pyrimidine phosphoribosyl transferases signature"  
2193...2252  
/note="Score 52: 20/22 ( 90%) matches, 0 gaps, 16 base loop"  
2343...2347  
2353...2349  
/gene="SCE66.03"  
/note="SCO3124"  
2353...2949  
/gene="SCE66.03"  
/note="SCE66.03, ribosomal L25p family protein, len: 198 aa; similar to many members of the ribosomal L25p family (includes ribosomal protein, general stress proteins and glutamyl-tRNA synthetases) eg. SW:P14194 (CTC\_BACSU) general stress protein from Bacillus subtilis (203 aa) fasta scores; opt: 327, z-score: 393.9, E(): 1.8e-14, 31.6% identity in 196 aa overlap. Contains Pfam match to entry PF01386 Ribosomal L25p, Ribosomal L25p family and prosite match to PS00017 ATP/GTP-binding site motif A (P-loop)."  
/codon\_start=1  
/transl\_table=11  
/product="ribosomal L25p family protein"  
/protein\_id="CAB95917.1"  
/db\_xref="GI:8994757"  
/db\_xref="SPTREMBL:Q9K3T9"  
/translation="MSEVKLTAEITRTFGKGAARRIRRNKVPQVLYGHGSDPLHLTL PHELLLAURTSNVLIDIDIKTNELAIKPSQVDPDKGFLEHVDQLVKRGETVSV EIPVQAEGLAPGGLLEIVLDALPVEAEATHIPQQTIVSVAGLEAGASIHAKDIALP SGVKLDVGDGTIVQLVLSAQAEAPAGEGEGEAGAA"  
2368...2625  
/gene="SCE66.03"  
/note="Pfam match to entry PF01386 Ribosomal L25p, Ribosomal L25p family, score 94.50, E-value 2.2e-24"  
2533...2556  
/gene="SCE66.03"  
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
3047...3052  
3055...3657  
/gene="SCE66.04"  
/note="SCO3125; pth"  
3055...3657  
/gene="SCE66.04"  
/note="SCE66.04, pth, peptidyl-tRNA hydrolase (EC 3.1.1.29), len: 200aa; strongly similar to many eg. SW:P23932 (PTH\_ECOLI) peptidyl-tRNA hydrolase from

Escherichia coli (194 aa) fasta scores; opt: 449, z-score: 529.8, E(): 4.9e-22, 41.1% identity in 185 aa overlap. Contains Pfam match to entry PF01195 Pept\_cRNA\_hydro, Peptidy1-cRNA hydrolase and Prosite match to PS01196 Peptidy1-cRNA hydrolase signature 2."

/transl\_table=1  
/product="peptidy1-cRNA hydrolase"  
/protein\_id="CAB95918.1"  
/db\_xref="GI:8894758"  
/db\_xref="SPTRMBL:09K3T8"  
/translation="MDVTTDAGAPMLVAGLGNPEPYASNRNNGFVADILAEIRGA  
AFKHGKAQGVNVEGRIGPPGPNRRVILAKPSEPMVSGVYVLRDVFYVGNVY  
AHDHLDIDYGVNLKLGGDNGHNGKSTYSGLADYHVRVFGIGRPDGMVADPFY  
LRDPSSTERKELDYFVDRADAVEALVIEGLERASAYNS"  
1085. .3651  
/gene="SC86.04"  
/note="Pfam match to entry PF01195 Pept\_cRNA\_hydro,

Query Match 3.0%; Score 45; DB 1; Length 28527;  
Best Local Similarity 48.3%; Pred. No. 14;  
Matches 126; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 824 CATTCCGGACTTCAGCAGAAAGTCCCGGATTCAACGGCTACAGCCGACGGA 883  
DB 11474 CTTTACCGCACTCTCTCGCCCGGATCCCGAACCTCCCGCGACCCGCCGATGCGG 11415  
QY 884 TACCCGCGAAGCAGCGGGAACTCTACCCAGCAGAAAGTGACAGATTCTTCGCGCCG 943  
DB 11414 ACCTCGGCTGCGCCCGCGCAACGTCACCGTCTGCTGCGCAGCCGTCGACCGCCG 11355  
QY 944 GCATGTCGCGCGCTTGGAAGAGCGCGCGGATTGACCTTGGCGCGAGCGGAGACCC 1003  
DB 11354 GCATACCGGCTACGACAACTGCGCCCGGATGCTCGAACGCGCCGCACTACCGCGCC 11295  
QY 1004 CAGCGCTCGCGCGCAGACCGCGAGAGAAACAATTGCAATTACTAGACCGCTGACG 1063  
DB 11294 CCACCTCGCGCGCGCGCGCACTGCACTTGGCGCGCGCGCGCGCTCTGAGACCCCG 11235  
QY 1064 ACAGAGCGCGCTGCGCCCTTGC 1084  
DB 11234 ACAGAGCGCTGCGCACTCTCTCG 11214

RESULT 12 SC3C9 20721 bp DNA linear BCT 12-MAY-2002  
LOCUS Streptomyces coelicolor cosmid 3c9.  
DEFINITION AL592262 AL645882  
ACCESSION AL592262.2 GI:20520790  
VERSION  
KEYWORDS AbA-like protein; dioxygenase; DNA-binding protein;  
DNA-methyltransferase; glutamate synthase small subunit; hydroxylase;  
integral membrane protein; putative metal-citrate protein;  
secreted hydrolase; sugar kinase; uvrB, exonuclease ABC subunit B.  
SOURCE Streptomyces coelicolor A3(2).  
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 20721)  
AUTHORS Redenbach, M., Kleser, H.M., Denapate, D., Bichner, A., Cullum, J.,  
Kinash, H. and Hopwood, D.A.  
TITLE A set of ordered cosmids and a detailed genetic and physical map  
for the 8 Mb Streptomyces coelicolor A3(2) chromosome  
JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)  
MEDLINE 97000351  
PUBMED 8843436  
REFERENCE 2 (bases 1 to 20721)  
AUTHORS Warren, T. and Harris, D.  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 20721)  
AUTHORS Cerdano, A.M., Parkhill, J., Barrell, B.G. and Randeram, M.A.  
TITLE Direct Submission  
JOURNAL Submitted (25-JUN-2001) Streptomyces coelicolor sequencing project.

Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrill@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK  
On May 9, 2002 this sequence version replaced gi:14571765.  
Notes:  
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics  
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web  
(URL: [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/)) CDS are numbered using the following system eg SC7B7.01c, SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).  
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.  
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(12):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, ggc, tgg or (atc)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 3c9 overlaps cosmid Stc54 and cosmid St7H2.  
Location/Qualifiers

FEATURES  
source  
1..20721  
/organism="Streptomyces coelicolor A3(2)"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
/clone="cosmid 3c9"  
complement(248..508)  
/note="Pfam match to entry PF00271 helicase\_C, Helicase conserved C-terminal domain, score 75.00, E-value 1.5e-18"  
complement(1460..1873)  
/note="Pfam match to entry PF00270 DEAD, DEAD/DEAH box helicase, score 27.00, E-value 1.2e-07"  
complement(11742..11765)  
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
complement(11946..2824)  
/gene="SC3C9.02c"  
/note="SC01967"  
complement(11946..2824)  
/gene="SC3C9.02c"  
/note="SC3C9.02c"  
/gene="SC3C9.02c"  
/note="SC3C9.02c, possible integral membrane protein, len: 292 aa; similar to TR-09K214 (EMBL:AL535863) Streptomyces coelicolor putative integral membrane protein SC68A.15c, 244 aa; fasta scores: opt: 703.2 z-score: 726.3 bits: 142.1 E(): 7.6e-33; 48.430% identity in 223 aa overlap. Contains possible hydrophobic membrane spanning regions"  
/codon\_start=1  
/transl\_table=1  
/product="putative integral membrane protein"  
/protein\_id="CAC42741.1"  
/db\_xref="GI:14571767"  
/db\_xref="SPTRMBL:09J3J4"  
/translation="MHSTVGGFCYGLVPLVAVYLMACAGLALGRTARAVLSNR  
PGMLAGSAATGSGITWTHFVAMGFTIHTPIRYDMITPFAVLAVIMVGVIPIV  
GVRGALFATGCTITGLGIAEMHYLGAAGHMLDQGLYNTFTVAASVIVAMATLA  
ALMAAGVGRGLMVGALITMGLAVTGMYTGAALVAVLSTGASVGGSPAEFLAP  
MLIGPLAFLLAGVVMFDELVMGRPAVPRARRKGPIDAHTEVPRTRIRHARRPR

RBS  
gene  
CDS  
misc\_feature  
misc\_feature  
misc\_feature  
gene  
CDS

TRHPLGHRRTTPQNR"  
3106..3110  
3116..3985  
/gene="SC3C9.03"  
/note="SC01968"  
3116..3985  
/gene="SC3C9.03"  
/note="SC3C9.03", probable secreted hydrolase, len: 299 aa;  
similar to SW:GIFQ\_BACSU (EMBL:Z26522) Bacillus subtilis  
glycerophosphoryl diester phosphodiesterase (EC 3.1.4.46)  
GlpO, 293 aa; fasta scores: opt: 495 Z-score: 576.9 bits;  
114.7 E(); 1.6e-24; 34.114% identity in 299 aa overlap.  
Contains possible  
peptide sequence"  
/codon\_start=1  
/transl\_table=11  
/product="putative secreted hydrolase"  
/protein\_id="CAC42742.1"  
/db\_xref="GI:14571768"  
/db\_xref="SPTREMBL:Q93J33"  
/translation="MHVRAVAVTTTALLGLVALTAPLSHARADEAGDDGPTTVARHGAS  
GYAPENTLAADRAAELGRVWENOVORTRGELVVLHDESLARTTVEVFDRSPW  
KQVDTAAETARDAGSNWFGSEYAGRVPTLEQYVHRVHNHKKLLELSPGLYFGI  
EQQTLYLANEGWLDNRVAGRLVQSFDSIRTVDLKPVKVTKGLTGPVASELPA  
YAEFADQINPSYSLMSYSAVHAFTGPHGSPLEVLVTWYNDATTARRVAGYDVGI  
ITNKPVDVRDAVD"  
4023..4577  
/gene="SC3C9.04"  
/note="SC01969"  
4023..4577  
/gene="SC3C9.04"  
/note="SC3C9.04", possible DNA-methyltransferase, len: 184  
aa; similar to SW:OGT MYCPA (EMBL:AJ011838) Mycobacterium  
paratuberculosis methylated-DNA protein-cysteine  
methyltransferase (EC 2.1.1.63) Opt, 165 aa; fasta scores:  
opt: 355 Z-score: 399.1 bits; 80.3 E(); 1.3e-14; 40.964%  
identity in 166 aa overlap. Contains Pfam matches to  
entries PF02870 Methyltransf\_1N, 6-O-methylguanine DNA  
methyltransferase, ribonuclease-like domain and PF01035  
Methyltransf\_1, 6-O-methylguanine DNA methyltransferase,  
DNA binding domain and match to Prosite entry PS00374  
Methylated-DNA--protein-cysteine methyltransferase active  
site"  
/codon\_start=1  
/transl\_table=11  
/product="putative DNA-methyltransferase"  
/protein\_id="CAC42743.1"  
/db\_xref="GI:14571769"  
/db\_xref="SPTREMBL:Q93J32"  
/translation="MDSHGQYQQVVTWVGTDIGPLLLAAATHDGLVNVVVFHATDTR  
GRALERAAALGTEPVEAYAGSPLLAEAIQVEAYFAGRRDFELPLDWSLISGFNRQV  
LRELASGVPYGVGVGDLAGRGPGAAQAVGMAMGANPLPVVPPCHRVVSDGGIG  
GFGGVDTKRLIALSGVLPPLP"  
4050..4298  
/gene="SC3C9.04"  
/note="PFam match to entry PF02870 Methyltransf\_1N,  
6-O-methylguanine DNA methyltransferase, ribonuclease-like  
domain, score 27.70, E-value 0.00028"  
4305..4574  
/gene="SC3C9.04"  
/note="PFam match to entry PF01035 Methyltransf\_1,  
6-O-methylguanine DNA methyltransferase, DNA binding  
domain, score 11.20, E-value 1.9e-32"  
4461..4481  
/gene="SC3C9.04"  
/note="PS00374 Methylated-DNA--protein-cysteine  
methyltransferase active site"  
complement(4602..5045)  
/gene="SC3C9.05c"  
/note="SC01970"  
complement(4602..5045)  
/gene="SC3C9.05c"  
/note="SC3C9.05c, possible dioxxygenase, len: 147 aa;

similar to TR:Q910T8 (EMBL:AE004682) Pseudomonas  
aeruginosa probable ring-cleaving dioxxygenase PA2546, 143  
aa; fasta scores: opt: 178 Z-score: 231.1 bits; 48.7 E();  
2.9e-05; 33.33% identity in 135 aa overlap. Contains Pfam  
match to entry PF00903 Glyoxalase, Glyoxalase/Bleomycin  
resistance protein/Dioxxygenase superfamily"  
/codon\_start=1  
/transl\_table=11  
/product="putative dioxxygenase"  
/protein\_id="CAC42744.1"  
/db\_xref="GI:14571770"  
/db\_xref="SPTREMBL:Q93J31"  
/translation="MSLGAMTDHTRLDHIVLWVRDPAAGVEKNLGLPLRITEX  
AACTVSPFVRUNDETIIDLAPHAMADRMNVPQADASAGHPVNHICVSLSPHDFDAL  
RTRLEQSIPIVSDLSYDYGARGMARRSFYFGDPDGNIIIRHYE"  
complement(4620..4991)  
/gene="SC3C9.05c"  
/note="PFam match to entry PF00903 Glyoxalase,  
Glyoxalase/Bleomycin resistance protein/Dioxxygenase  
superfamily, score 46.90, E-value 2.9e-12"  
5155..5161  
5172..6077  
/gene="SC3C9.06"  
/note="SC01971"  
5172..6077  
/gene="SC3C9.06"  
/note="SC3C9.06, conserved hypothetical protein, len: 301  
Query Match 3.0%; Score 44.4; DB 1; Length 20721;  
Best Local Similarity 55.1%; Pred. No. 18;  
Matches 87; Conservative 0; Mismatches 71; Indels 0; Gaps 0;  
QY 839 AGCAGAGGTCCCGAATTCACGCGTACACGTCGCCAACGCGATACCCGCGAACAGCG 898  
Db 5866 AGCAGTCTGATCCCGGACTCACGCGCGTGTCTCGCGAGCGGTGCACGCTCGCGAGG 5925  
QY 899 CGGGAACCTCACCACGAGAGGTGCAGAGTCTCTCGGCGCGGATGTCGCCGTT 958  
Db 5926 CGGAGGTGTTCAGGCGGCGAGCGGTTCACGCGGTCTCTCTCGAACGCGCTGGTGGCTCA 5985  
QY 959 GGAAGAGCGCGCGCGATGACCTCGCGCGCGAGCGCG 996  
Db 5986 CCGACGCGCGCTCGCTGAGCGCAACCTGGCGCGGTG 6023  
RESULT 13  
LOCUS AE010452 11738 bp DNA linear BCT 01-AUG-2002  
DEFINITION Methanopyrus kandleri AV19 section 151 of 157 of the complete  
genome.  
ACCESSION AE010452 AE009439  
VERSION AE010452.1 GI:19888342  
KEYWORDS  
SOURCE Methanopyrus kandleri AV19.  
ORGANISM Methanopyrus kandleri AV19.  
Archaea; Euryarchaeota; Methanopyri; Methanopyrales;  
Methanopyraceae; Methanopyrus.  
REFERENCE 1 (bases 1 to 11738)  
AUTHORS Shcherbinina, O.V., Shakhova, K.V., Makarova, K.S., Polushin, N.N.,  
Natalie, D.A., Rogozin, I.B., Tatusov, R.L., Wolf, Y.I., Stetter, K.O.,  
Malykh, A.G., Koonin, E.V. and Kozyavkin, S.A.  
The Complete Genome of the Hyperthermophile Methanopyrus kandleri  
AV19 and Monophyly of Archaeal Methanogens  
Unpublished  
JOURNAL 2 (bases 1 to 11738)  
REFERENCE 2 Slesarev, A.I., Mezhevaya, K.V., Makarova, K.S., Polushin, N.N.,  
Shcherbinina, O.V., Shakhova, K.V., Belova, G.I., Aravind, L.,  
Natalie, D.A., Rogozin, I.B., Tatusov, R.L., Wolf, Y.I., Stetter, K.O.,  
Malykh, A.G., Koonin, E.V. and Kozyavkin, S.A.  
Direct Submission  
JOURNAL Submitted (04-FEB-2002) Fidelity Systems, Inc., Gaithersburg, MD  
20879

FEATURES		Location/Qualifiers
source	1..11738	/organism="Methanopyrus kandleri AV19"
gene	122..1183	/db_xref="taxon:190192"
CDS	122..1183	/gene="MK1628"
gene	/codon_start=1	
	/transl_table=11	
	/product="Uncharacterized protein"	
	/protein_id="AA02841.1"	
gene	/db_xref="GI:1988343"	
	/translacion="MAERIVREVLREARRGVSDLPFSGEGVNEITDRISRVLGS	
	VGASEDVRRESILEETSSKLPPTYLSSIAETVSOVLPMGEEDADALPDHDFSEKYL	
	DHLLEGAEFVCLASPVVSSPKDLVBEGRVSRIGLSDRLDELXILLVAENREEVILE	
gene	MAISGFVREARRRREBDLGHCKVYANRELAAGAKNTVTSLSRLAAKEVYITNP	
	KADDEVGNANTYEQIVSEFNSQSEAKQRFDEGKPAITFVRWDDRTFVAVTR	
	ENGERWTFVELEDHGFVNEREGLPVRGVRVHRFDLRGSRRRGQDRLSLRTOA	
	REDPSGASGCRAGREREMER"	
CDS	1167..1460	
gene	/gene="MK1629"	
	1167..1460	
	/gene="MK1629"	
	/codon_start=1	
gene	/transl_table=11	
	/product="Uncharacterized protein"	
	/protein_id="AA02842.1"	
	/db_xref="GI:1988344"	
gene	/translacion="MRWKNDAKKYKQITGRPYDGLVVMGLDPPDRSDPYVAGSKLL	
	VIRCESGSGITVEFVDVADPGNLPWVSQSVLDGLVLPDRSGEGTPTLTFVYD"	
	complement(1485..2831)	
	/gene="MK1630"	
CDS	complement(1485..2831)	
	/gene="MK1630"	
	/codon_start=1	
	/transl_table=11	
gene	/product="tRNA/tRNA cytosine-C5-methylase"	
	/protein_id="AA02843.1"	
	/db_xref="GI:1988345"	
	/translacion="MENQIALVAKALVYENRGNPIPEAVERTCRDRPIDVRSQVAPY	
gene	YETIRRLDELVAAGSDAHPEQVRSPPYRQILRGITLKKIKNPPAPATVDCVAPY	
	AKRLVGSAGAFVAVLVGVRVSVQDVEDRWTERRLAKYGHPEVFLVLDLFEQ	
	DRSRVELLRANRNPQPYLRINRLKLDPLVADGVLEERCAVGPTEPTLBEVRALVRG	
	RGYGRARWREGLPVODLASASAAALSAPCEVLDVCAAGSRTHTTAPREMLDEGE	
gene	VMAVDRSEMGARVLERRCRLGITCVATTICDRARGLTVDLPVDPDRITLVPPSGTIG	
	VNNRNPDSRMKPKPLERFAEROMEILEPARIAEHGGTLVYVSTCSVSEENREAIIVKR	
	ALEREDVLYNDAGVLSPGIEFRGERFSYKRYRNPFRIDTAGFVAAKKGKG"	
	2858..3598	
CDS	/gene="arg8"	
	/note="MK1631"	
	2858..3598	
	/gene="arg8"	
gene	/codon_start=1	
	/transl_table=11	
	/product="Acetylglutamate kinase"	
	/protein_id="AA02844.1"	
gene	/db_xref="GI:1988346"	
	/translacion="MEVIKVGGEVLDVRVEDLARVIDSLIVHGGGEPSVDYMERKGLE	
	PRFVRGLVTDRETLOVMMVLAGLVNKRVLVABSEGJNALGSLGVGGILLIEKRS	
	EVVDGEVLDVGVDRVNAELLESLLDYGVPVVAPLGAGEDGVVNNVADPAGA	
gene	IAGAVRADRLVLTIDVPGVLEDDPETTLIERVVRPDEVLEEKGIVTGVMPKLEMA	
	KMAVAGCREAVITNLEGLLEGRTIVR"	
	3532..3846	
	/gene="MK1632"	
CDS	3532..3846	
	/gene="MK1632"	
	/codon_start=1	
	/transl_table=11	
gene	/product="Uncharacterized protein conserved in archaea"	
	/protein_id="AA02845.1"	
	/db_xref="GI:1988347"	

---

gene	/translacion="MPRGDRHGRGVARGEGCHREVAALHGEKIEKYISKLPSTVINEC	
	KPRDRAGERLLDLASRYFPDARYFLDRGEMVYAFCTLSYAAALLRAGAEVGVLDVPEQ	
	EV"	
	3933..4556	
CDS	/gene="MK1633"	
	3933..4556	
	/gene="MK1633"	
	/codon_start=1	
gene	/transl_table=11	
	/product="Uncharacterized protein"	
	/protein_id="AA02846.1"	
	/db_xref="GI:1988348"	
gene	/translacion="MVRPVSVGHGGSVEVLHPSDLPPOLTSLOCKIKINGARVISRR	
	SNIVAFRGRSLPLNDLDGFMSSIAARESEVVMGDPYLRLVLAVEGERIYVNTDRLL	
	ACLPFETIELKNEBITVECEGPDIVLACSSPVVASLKLEDTXHVAGSILCCMMNI	
	HTXWVGFLLTGCGHMLSTTKKPAATEKAGSTKTTAPRVSSWI"	
CDS	complement(4615..4690)	
	/product="tRNA-Gly	
	/note="anticodon:GCC"	
	complement(4725..5360)	
gene	/gene="MK1634"	
	complement(4725..5360)	
	/gene="MK1634"	
	/codon_start=1	
CDS	/transl_table=11	
	/product="Zn-dependent hydrolase"	
	/protein_id="AA02847.1"	
	/db_xref="GI:1988349"	
gene	/translacion="MRVFRVGGKGITPSPDCNCTLLAVGDEGILLDLGASGEVIDRLP	
	GNVUVRYALLTHSHFDHAAAGPDALGAEVGVHAAEVAVEREGDRLISAAYLDRPM	
	PAYSPFTPFDEGTFDVGEVEVLYTPHSGSCCPFLVGLDAPFGDVFPGCGRMD	
	LPGDRKKLCESSERLLSTGVRSIFPGHGEVYIGAVPAIEALAREAKOI"	
CDS	5403..6257	
	/gene="MK1635"	
	5403..6257	
	/gene="MK1635"	
gene	/codon_start=1	
	/transl_table=11	
	/product="N6-adenine-specific DNA methylase"	
	/protein_id="AA02848.1"	
gene	/db_xref="GI:1988350"	
	/translacion="MRPEELALVAREALRRVPPDPDANTHTVVSDDALHVLITNO	
	RKNVEGCGKRNROYLEGVACGYLAAGHGVREAVSVVCGADETEREPRVPLTRALK	
	EGVIEBGSPIHRCVERIGIGVVLGARELVGNPRLVEVEBELGEERAVDPFTGTGQ	
gene	ALAAVEAGFEQVAIVDIRVAFREERLSEEGVEIEADFRVDLREFFPIDLLTADPP	
	VASTLEFPEKSEERRPVDVAVVCHGFSWTRAVREINRGFLIELFEDVEPVSXVGHSL	
	SVCRILD"	
	complement(6596..7210)	
CDS	/gene="MK1636"	
	complement(6596..7210)	
	/gene="MK1636"	
	/codon_start=1	
gene	/transl_table=11	
	/product="Uncharacterized protein specific for M.kandleri, MK-1 family"	
	/protein_id="AA02849.1"	
	/db_xref="GI:1988351"	
gene	/translacion="MVRSHAPIYNPVERIRALRLVREYHRRRKPSPLEVTYRVNGS	
	TGQPYVAARMRDRPKHGRITLVLGPEMBVSFVEMVLDRFVLALHAKNLR	
	SVLKTLLTEVSDLPYKARVLLRGJALAFDARPESEPIRDLRELPDRLESFAVTR	
	LGGMPAHYSYLKRVYIHRKGSIDEGHEVDVGLFEGRWKLDQRG"	
CDS	complement(7414..8076)	
	/gene="MK1637"	
	complement(7414..8076)	
	/gene="MK1637"	
gene	/codon_start=1	
	/transl_table=11	
	/product="Orphan DOD family homing endonuclease"	
	/protein_id="AA02850.1"	
CDS	/db_xref="GI:1988352"	
	/translacion="MTSSPYGSEEGRLRWLEDDRLALGSPYRYEYGDRAVEAVHSE	
	GFLALIGLMPGCRKGRVHVGIVRQDPDVALPEFLSGLPAGACWISABENVEYG	
	IAQASPMGSESEPLEGVSSLLKKTCTTVYARVLEAGSYETISVDGRASVGLGFRHEHYE	







misc\_feature /function="approximate span of module 3'"  
97. .1482  
/gene="eryA"  
/note="putative"  
/function="approximate span of beta-ketoacyl ACP synthase  
of module 3'"  
1693. .2670  
/gene="eryA"  
/note="putative"  
/function="approximate span of acyltransferase domain of  
module 3'"  
3406. .3921  
/gene="eryA"  
/note="putative"  
/function="approximate span of beta-ketoreductase domain  
of module 3, possibly non-functional'"  
4171. .4428  
/gene="eryA"  
/note="putative"  
/function="approximate span of acyl carrier domain of  
module 3'"  
4471. .10722  
/gene="eryA"  
/note="putative"  
/function="approximate span of acyl carrier domain of  
module 3'"  
4471. .5847  
/gene="eryA"  
/note="putative"  
/function="approximate span of beta-ketoacyl ACP synthase  
domain of module 4'"  
6054. .7026  
/gene="eryA"  
/note="putative"  
/function="approximate span of acyltransferase domain of  
module 4'"  
7165. .9216  
/gene="eryA"  
/note="putative"  
/function="approximate span of dehydratase and  
enoylreductase domains of module 4'"  
9433. .9984  
/gene="eryA"  
/note="putative"  
/function="approximate span of beta-ketoreductase domain  
of module 4'"  
10225. .10483  
/gene="eryA"  
/note="putative"  
/function="approximate span of acyl carrier domain of  
module 4'"  
10723. .20235  
/gene="eryA"  
/note="ORF3 encoding modules 5 & 6 for  
6-deoxyerythronolide B formation; putative"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AA026495.1"  
/db\_xref="GI:152695"  
/catalytic="MSGNGKTEBKRYIKTIVTLDSTVRLAEVEHRAAGEPIAIV  
GNAKCFPGDVDSFESFVSGGDAIAEPADRGWPPDPRLAGMTLAAADPAGF  
FGISPRELAMDPOORIMLEISWALERADHGRVATGVPTFGVIGRTDGFPLAT  
APDEVLTGVTGTAASVAGSVACGLDEGPATVTTCASSGTLHLHLMESLRDEC  
GLALAGVTVMSPGAFTEPRSGGLAEDCRKPFASKADFGLAEGAGVLIORLSA  
ARREGRPVLAIRGSAYNODGASNGLTAPSGAOCQVIRALENAGVAGADVYEAH  
GTCGTGDPTEVHALISTYGAERDPDPLMTGSKNIGHTQAAAGVAGVAKVIALR  
HGEHRTIHPDEPRPOTIFMDLGAISVVSQSRPAEERRRRGVSFGISGTHAVIV  
BEAPDEPRPAPDSGPVPLVSGRDEQARPAAGRLAHLRAEPNSLRDTGFLAT  
RRSAEHRVAUVGGRDALAGRLAVADGRIADTTATGAGRTRGVAMWPGGAGWOG  
NARDLRESQVAFDSIRDCERALA.PHYDWELTDLTSGARPLDRVAVPALFAVWVSL  
AALMRSGVEPAAVVGSHQGEIAAAHVAGALLLEDAKLVAVSRVLARLGGCGMAS  
FGLGTGQAERIGRFAGALSTIASVNGRSVVAGSGSPDELDIEACGACGATARIPIV  
DVASHSPVESIRELITELTAGISPVASADVALYSTTGGQIDPATDTGYWYANRLO  
VRFQATQRLAEGDAFVBSHPVLTGTEITLSDALPADAGCVVGTTLARDNGCI

ADPHTALGFAVAGVEMDSAPADAPVLELVPEFORRWYLPITPGRARDDDDMM  
RYOVNREABEWESASLAGRVLYTGGVSELSDDAIRGLOSQATVITCVBSRSTI  
GTALAEADDTALSTVLSURDGEAVDPSLDLALVQLAGAERAPLMLITRAVQV  
ADGELVPPQAMVAGLGRVVGIEOPGRMGVLDVLDADAASITSLAAVLADPREEVY  
AIRADGIVKARVLPAPARAARTMSPRGWGLVTGTGGIGAMVAMWLARSGAEHVLV  
GRGADAPGASELREELITAGTCVITIAACDVADRRLAVILAERAEGRVTASVWMAA  
GVSTSTPLDILTAEFTEIADVYRGCTVNLDELCDPLAFVLFSSNAGVMSPGIASY  
AAANAFLDGFARRRSEGAAPVTSIANGLAQNMGMDEGGELRYEQGRAMPDRAVE  
ELHTLDHGCOTSVSYVDMDRRPVELFTARRRPLPDELAGARARAROSEEGPLOR  
LAASTBRREBHIAHLIAEVAAVVLGHGDDAIDDRAPAFDUGPRTSMYAVULNRILAA  
VTGRENAVIVFHPITITRLADHYLERLILGALEAEOAPLVAVEFKDDADDTLAVGMA  
CRFPFGVANHPEGLMEFTVGGADVATEMPDRMDLALFDPPORHRTSYSRHCAFVLD  
GAADFDAFPGIISPREALAMDPOORQVLENTLEENAGIPEHSLRGSDFVFLGAAY  
QGYODAVPEDESEGLITNSAVVSGVAVALVGLIEGAAVYVDTACSSSLVALHSAAC  
GSLRDGCGLAIVAGVSVMAPEVFTEFRQGLAVDGCRCFASEADPFGFAGVAV  
VLIORLDBARAGROYIVAGVAGSAINODASNGLAAPGVAOORVIRAMRAGITGA  
DVAVEFAHGTCTPLGDPVESAALLATYSGRSRSGPVLIGSKVSNIGKQAAVAGVY  
IKVULGNRLGVPPMLCRGERSPLIEMSSGUYELNAPSMPAPADSVIRARAGVAFVY  
SGTHAVITIAPEPPEPLPEPGVGVLAHANSVPVLSARTETTLAAQRLAESAVD  
SVPLTALASALATGRALHPRRALLAGDHEOLRGULRAVEGVAAPATTGTASAGVY  
VFPPGQAQMGMARGLSVPFABSIACPAVISEVAGFASVLEFORDAELER  
VDVQPVLFESVMSLARLWGCAGVSPSAVIGHQSEBIAAAVAGVLSLEDCRVALLR  
AKALRALAGKGMVSLAPEGARALIAWEDRIEVAANSPSSVSVGDEPALVALR  
ARCEDEGVAKTLPVDVYASHSRHEIRIETIADLDSISARRAALPLSTYHGERDQ  
ADMGRPYWYNLSOVARFDEAVAAVADDAHATFVMSHPHPLTAAVOBIADAVATGS  
LHRDTAEHILIAELARAHVAGVAVDRNRFPAAPRYALPNYFEPQRVLAPEVSDOL  
ADSRVDMRPLATTPVDLGGFLVHGSAPBSLTSVEGAGRVFVASARBASAL  
REVPEVAGVLSYHTGAHTLALHOSLGAERAPALMLTSSAVALGSESPEDQAM  
VWGIGRWGLETBERMCGVLDLPABEPAPDGBFAVACADADHEOYVALRIDHARYGR

Query Match 2.9%; Score 43.4; DB 1; Length 20235;  
Best Local Similarity 46.1%; Pred. No. 31;  
Matches 182; Conservative 0; Mismatches 211; Indels 2; Gaps 1;

Qy 845 AGGTCGCCGAAATTCAAACGCTACACGCTCGCCAAAGCGATACCCGGGAACAGCGCGGAA 904  
Db 12623 AGCCCGCCCGGCTGTCCGCCACTCGCAGGCGGATCGCCCGCCGCAAGTCCGCGG 12682

Qy 905 ACCTCACCCAGAGAGAAGTGCAGAGTTCCTCGGCGCGCCGATGTTCCGCGCTTGAMA 964  
Db 12683 CGGTACCTCGAGAGACGCCGCAAGCTGTCGGGTCGCGAGCGGCTCTTGAGCCGCG 12742

Qy 965 GCGCGCGCGGATTGACTTCGCGCCGACGCGGACGACCCAGCGCTCGCGCCGACCCG 1024  
Db 12743 TCCGCGCGCGGCAATGCGTCTTCGGGCTGGGCGACCGAGCGCGCGGAAAGGA 12802

Qy 1025 AGAGAAACAACAAATTTTGATTATCTGACACAGCTGACGAGGACGCGGCTGCCCTTG 1084  
Db 12803 TCGGGCGCTTCGCGGCGCGGCTCTTCATCGCTGATCAACGCGCCGCGCTGCTG 12862

Qy 1085 GTGCACACGTGAGAAAGCAACCTTCACAGACCTGGGCTGACCGCTGCACACTT-- 1142  
Db 12863 TCCCGGGGAGAGCGGCGCGCTGACGAGCTGATCGCGAGTGGAGAGCGGAGGATTA 12922

Qy 1143 CCAAGCTATGCGGTCCAGTATCCGTAAGGCCGAAACGCTGATCAGAACTTGCGTGC 1202  
Db 12923 CGGCGGCGCGGCACTCCCGTCACTACGCGCTCCACTACCGGAGGTGAGATCCGTCG 12982

Qy 1203 GGGCGTACTCGGCAAGACCGGCTCTTTTGG 1237  
Db 12983 AGGAGCTGCTGACCGAGCTGGGCGGCACTCTCCCG 13017

RESULT 15  
AR049368  
LOCUS AR049368 20235 bp DNA linear PAT 29-SEP-1999  
DEFINITION Sequence 3 from patent US 5824513.  
ACCESSION AR049368  
VERSION AR049368.1 GI:6005407  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
Unclassified.

REFERENCE 1 (bases 1 to 20235)  
AUTHORS Katz, L., Donadio, S. and McAlpine, J. B.  
TITLE Recombinant DNA method for producing erythromycin analogs  
JOURNAL Patent: US 5824513-A 3 20-OCT-1998;  
FEATURES Location/Qualifiers  
source 1..20235  
BASE COUNT 2534 a 7194 c 7767 g 2740 t  
ORIGIN

Query Match 2.9%; Score 43.4; DB 6; Length 20235;  
Best Local Similarity 46.1%; Pred. No. 31;  
Matches 182; Conservative 0; Mismatches 211; Indels 2; Gaps 1;

QY 845 AGGTCCCGCAATTCAACGGGTACACGCTGCCCAACCGCATACCGCGAACACAGCGCGGAA 904  
DB 12623 AGCCCGCGCGTCTGCGGCACTTCGCGGGCGAGATCGCCGCGCGACGTCGCGCGCG 12682

QY 905 ACCTCACCCAGCAGAGGTGACAGTTCTCGGCGCGCATGTTTCGGCCGTTGGAAGA 964  
DB 12683 CGCTCACCTGGAGAGCGCGCCAGCTCGTCGGTCCGGAGCCGGGTCTTGGCCCGGC 12742

QY 965 GCGGCGCGCCGATTGACCTCGGCGGACGCGGACACCCAGCGCTCGGCGCGCGACCCGC 1024  
DB 12743 TCGGCGCGCAGGCGGATGGCGTCTTCGGGCTGGGACCCGAGCAGCGCGCGCAACGGA 12802

QY 1025 AGAGGAACAATTTTCGATTACTCAGACACGCTGACGACGAGACGCGCTGCCCCCTTCG 1084  
DB 12803 TCGGCGGCTTCGCGGGCGGCTCTCCATCGCTCGGTCAACGGCCCCCGGTGCTCG 12862

QY 1085 GTGCACAGTGAGGAAGACGAAACCTCGACAGACCTGGGTGGACCGGTTCGACACCTT-- 1142  
DB 12863 TCGCGGGGAGAGCGGCGCGTGGACGAGTGATCGCCGAGTCCGAGGCCGAAGGCATAA 12922

QY 1143 CCACGCTATGGGTCCAGTATCCCGTACGGCCCGAAGACGTCGTGATGAGAACTTTCGCTC 1202  
DB 12923 CGGCGCGCGATCCCGTCACTACGCTCCCACTCACCGCAGGTGGAGTCGCTGCGCG 12982

QY 1203 GGGCGTGACTCGGCAAGACCGGCTCTCTTTTCG 1237  
DB 12983 AGGAGCTGCTGACCGAGCTGGCGGGCATCTCCCGG 13017

Search completed: December 4, 2002, 22:48:18  
Job time : 3356 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2002, 14:49:27 ; Search time 255 Seconds  
(Without alignments)  
13194.061 Million cell updates/sec

Title: US-09-926-084-8

Perfect score: 1494  
Sequence: 1 atgcgcctgcgcgtcgttgc.....ctacgcgcactacgcgtcgt 1494

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

N Geneseq\_101002:\*

1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*

2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*

3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*

4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*

5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*

6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*

7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*

8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*

9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*

10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*

11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*

12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*

13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*

14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*

15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*

16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*

17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*

18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*

19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*

20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*

21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*

22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*

23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*

24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1494	100.0	1494	21	AAA92280
2	43.4	2.9	29879	14	AAQ46806
3	42.4	2.8	9757	16	AAQ97686
4	42.4	2.8	9759	18	AAT98642
5	41.8	2.8	2433	22	AAH8043
6	41.4	2.8	1572	23	AAH52083
7	41.4	2.8	1572	23	ABLO9011
8	41.4	2.8	4186	23	ABLO9010
9	41.2	2.8	840	24	ABK95664

#### ALIGNMENTS

10	41.2	2.8	861	24	ABK95652	Grass cDNA for all
11	41.2	2.8	861	24	ABK95653	Grass cDNA for all
12	41.2	2.8	861	24	ABK95654	Grass cDNA for all
13	41.2	2.8	861	24	ABK95657	Grass cDNA for all
14	41.2	2.8	861	24	ABK95658	Grass cDNA for all
15	41.2	2.8	861	24	ABK95659	Grass cDNA for all
16	41.2	2.8	861	24	ABK95660	Grass cDNA for all
17	41.2	2.8	861	24	ABK95661	Grass cDNA for all
18	40.8	2.7	9759	19	AAV34766	Rubella virus RA27
19	40.8	2.7	9752	21	AAZ48249	Rubella virus Cend
20	40.6	2.7	1350	24	ABA99445	Actinoplanes sp DN
21	40.6	2.7	38064	24	ABA99465	Actinoplanes sp SE
22	40.2	2.7	1441	21	AAZ87284	S. venezuelae deso
23	40.2	2.7	13613	21	AAZ87319	S. venezuelae deso
24	40.2	2.7	13613	24	AAZ873043	Streptomyces venez
25	40	2.7	1503	20	AAV12539	Thermus flavus amy
26	39.6	2.7	861	24	ABK95651	Grass allerigen P 5
27	39.2	2.6	1584	22	AAH51986	Mycobacterium tube
28	38	2.5	467	24	ABK62208	Rat sequence diffe
29	38	2.5	515	24	ABK62166	Rat sequence diffe
30	38	2.5	861	24	ABK95656	Grass cDNA for all
31	38	2.5	861	24	ABK95662	Grass cDNA for all
32	38	2.5	861	24	ABK95663	Grass cDNA for all
33	38	2.5	2455	24	ABK63479	Grass cDNA for all
34	37.6	2.5	103519	22	AAZ08693	Microspora DNA
35	37.2	2.5	286	24	ABL75855	Corn tassell-derive
36	37	2.5	65140	22	AAZ17184	Streptomyces nous
37	37	2.5	125401	22	AAZ17186	Streptomyces nous
38	36.8	2.5	1248	21	AAZ67286	S. venezuelae deso
39	36.8	2.5	1248	24	AAZ67286	S. venezuelae deso
40	36.8	2.5	5970	21	AAZ5635	Streptomyces venez
41	36.8	2.5	5970	21	AAZ56003	Nucleotide sequenc
42	36.8	2.5	11878	22	ABA21379	Contig 002 from co
43	36.8	2.5	11999	22	ABA21380	Human nervous syst
44	36.6	2.4	68750	21	AAZ55867	Human nervous syst
45	36.4	2.4	861	24	ABK95655	Soratingu cellulos
						Grass cDNA for all

#### RESULT 1

AAA92280 standard; DNA; 1494 BP.

XX AAA92280;

DT 09-JAN-2001 (first entry)

DE Geotrichum candidum Dec 1 (FERM BP-7033) Dye DNA SEQ ID NO:8.

XX Geotrichum candidum Dec 1 (FERM BP-7033); Dye; decolouring activity;  
KW dye; peroxidase; dye-decomposition; waste; textile; ds.

XX Galactomyces geotrichum.

OS Galactomyces geotrichum.

XX Galactomyces geotrichum.

PN WO200050582-A1.

FD 31-AUG-2000.

XX 25-FEB-2000; 2000WO-JP01093.

XX 26-FEB-1999; 99JP-0050562.

XX (MEIJ) MEIJI SEIKA KAISHA LTD.

XX Syoda M, Sugano Y, Kubota H;

XX WPI; 2000-549408/50.

XX P-FSD; AAB23718.

XX Geotrichum candidum-originated peroxidase with high dye-decomposing  
PT activity on wide range of dyes, useful for waste treatment in e.g.



```
FT FT /*tag= C
XX XX /note= "orf 3"
PN MO931363-A.
XX 22-JUL-1993.
PD 17-JAN-1992; 92WO-US00427.
PF 17-JAN-1992; 92WO-US00427.
PR 17-JAN-1992; 92WO-US00427.
XX (ABBO ) ABBOTT LAB.
PA
XX Donadio S, Katz L, McAlpine JB,
PI
XX WPI; 1993-242804/30.
DR P-PSDB; AAR44430-32.
DR
XX
XX
XX Biosynthesis of specific polyketide analogues esp. erythromycin
PT cpds. - by introducing altered biosynthetic gene-contg. DNA into
PT microorganisms
PS
XX Claim 27; Fig 2; 133pp; English.
XX
XX This sequence represents a fragment of the Saccharopolyspora erythraea
CC genome, designated eryA. The polypeptides encoded by this region
CC are involved in the biosynthesis of the polyketide segment of
CC erythromycin. eryA is organised in modules and each module takes care
CC of one condensation step. The precise succession of elongation steps
CC is dictated by the genetic order of the modules. This fragment may be
CC specifically altered such that novel polyketide molecules of desired
CC structure are produced. Three types of alteration may be produced:
CC those inactivating a single function in a module which does not arrest
CC acyl chain growth; those inactivating a single function in a module
CC which does affect chain growth; and those affecting an entire module.
CC The mutations may be introduced by gene replacement.
XX
XX Sequence 29879 BP; 3672 A; 10596 C; 11545 G; 4066 T; 0 other;
SQ
```

```
Query Match 2.9%; Score 43.4; DB 14; Length 29879;
Best Local Similarity 46.1%; Pred. No. 0.13;
Matches 182; Conservative 0; Mismatches 211; Indels 2; Gaps 1;
```

```
QY 845 AGGTCCCGGAATTCAACGCGTACACGCTCGCAAACGGATACCGCGGAACAGCGCGGAA 904
DB 22822 AGCCCGCGCGGTCTCTCGCCACTCGCAGGCGGATCGCCCGCGCACTGCGCGCGG 22881
QY 905 ACCCTACCCAGCAGAGAGATGATGATCTCTCGGCGCGCGCATGTTCCGCGCGTGAAGA 964
DB 22882 CGCTACCTTGAAGAGACGCGCAAGCTCTGCGCGGTCCGAGCGCGGTCTGAGCCCGG 22941
QY 965 GCGGCGCGCGATGACTCTGCGCGCGAGCGAGCAACCAAGCGCTGCGCGCGGACCGGC 1024
DB 22942 TCGCGCGCGCGAGCGCGATGCGTCTGCGCGGTCCGAGCGCGCGCAAGCGCA 23001
QY 1025 AGAGGAACAACAATTTCATTCAGACAGCTGAGAGCGAGAGCGCGTGGCCCTTGG 1084
DB 23002 TCGGCGCGTTCGCGCGCGCGCTCTCATCGCTCTGATCAACGCGCGCGGTCTG 23061
QY 1085 GTGCAACGCTGAGAGAGCAACCGCTCGACAGACTCTGAGTGAACCGGTGACACTT-- 1142
DB 23062 TCGCGCGCGAGAGCGCGCGCGTGAAGAGCTGATCGCGGAGTCAAGGCGCAAGGATTA 23121
QY 1143 CCAAGCTATGCGGTCCAGATCCGCTACGCGCGCAAGAAAGCTGATGACAAATTCGCTG 1202
DB 23122 CGGCGCGCGCGCATCCCGCTCACTACGCTCCCACTACCGCAGGTGAGTGCCTCGCG 23181
QY 1203 GGGCGTGAATTCGCGCAAGACCGCGGTCTTTTTCG 1237
DB 23182 AGGAGCTGTACGACGAGCTGAGCGGAGCATCTCCCG 23216
```

RESULT 3

```
AA097686
ID AA097686 standard; RNA; 9757 BP.
XX
XX AA097686;
AC
XX 27-FEB-1996 (first entry)
DT
XX
XX Infectious rubella virus RNA.
DE
XX
XX Rubella; vaccine; mutant; epitope; virus; autoimmune disease;
KM pregnancy; foetal infection; vector; plasmid; ss.
XX
XX Rubella virus.
OS
XX
XX Key Location/Qualifiers
FH 41..6658
FT CDS
FT /*tag= a
FT /*product= N-terminal transcript.
FT misc_difference 2261..2263
FT /*tag= b
FT /*transl_except= CGU encodes Ala.
FT misc_difference 6605..6607
FT /*tag= c
FT /*transl_except= GUC encodes asparagine or aspartic acid.
FT misc_difference 8460..8462
FT /*tag= d
FT /*transl_except= CUG encodes Proline.
FT misc_difference 8463..8465
FT /*tag= e
FT /*transl_except= CUC encodes Cysteine.
FT misc_difference 9075..9077
FT /*tag= f
FT /*transl_except= UGG encodes Methionine.
XX
XX US5439814-A.
XX
XX 08-AUG-1995.
XX
XX 28-JUN-1991; 91US-0722334.
XX
XX 28-JUN-1991; 91US-0722334.
XX
XX 19-JUL-1993; 93US-0093453.
XX
XX (GEOR-) GEORGIA STATE RES FOUND INC.
XX
XX Dominguez G, Frey TK, Wang C;
XX
XX WPI; 1995-283097/37.
XX
XX P-PSDB; AAR79048; AAR79049.
XX
XX New DNA encoding infectious rubella virus - esp. non-pathogenic
PT mutant virus for use in vaccines having reduced side effects
XX
XX Claim 2; Columns 9-18; 27pp; English.
XX
XX Non-pathogenic mutants of the DNA corresponding to this sequence are
CC useful in vaccines (which may include epitopes from other viruses).
CC The mutant vaccines are less likely to cause foetal infections,
CC autoimmune disease or neurological symptoms, so can be administered
CC safely to older or pregnant women. The mutant sequence is
CC preferably present in a vector, especially a bacterial plasmid that
CC allows replication of the sequence.
XX
XX Sequence 9757 BP; 1458 A; 3782 C; 3008 G; 1509 U; 0 other;
SQ
```

```
Query Match 2.8%; Score 42.4; DB 16; Length 9757;
Best Local Similarity 47.8%; Pred. No. 0.15;
Matches 86; Conservative 8; Mismatches 86; Indels 0; Gaps 0;
```

```
QY 841 CAGAAAGTCCCGGAATTCAACGCGTACACGCTCGCAAACGGATACCGCGGAACGCGCG 900
DB 5600 CAUACGAGCCCGGAGUGCAUGCGUGGAGAGGCCCAUACACCAACGCAUCCAG 5659
```

QY 901 GGAACCTTACCCAGCAGGAAGGTGCAGAGTTCTTCGGCGCGGCATGTTTCGGCGCGTGG 960  
Db 5660 GUCCAGUUCACUGAGUUCACUAGAACACAGACCCUCCUACUGCGGACGUCGAGCUCGAG 5719  
QY 961 AAGAGCGCGCGCGATTGACCTTCGCCCGCAGCGGCGAGACCCAGCGTTCGGCGCGGAC 1020  
Db 5720 AUUAGCGCGCUCUCUUGGCGUCCUUGCGCGGAGACUACCGCGCGUCCGCGCGCGC 5779

RESULT 4  
AAT89642  
ID AAT89642 standard; cDNA; 9759 BP.  
XX  
AC AAT89642;  
XX  
DT 26-JAN-1998 (first entry)  
XX  
DE Infectious rubella virus cDNA clone.  
XX  
KW Rubella: German measles; foetal death; birth defect; vaccine;  
KW immunisation; non-pathogenic mutant; genetic engineering; ss.  
XX  
OS Rubella virus.  
XX  
FH Key Location/Qualifiers  
FT CDS 41..6388  
FT FT /\*tag= a  
FT FT /note= "encodes p150 and p90"  
FT CDS 6509..9700  
FT FT /\*tag= b  
FT FT /note= "encodes capsid, envelope 1 and envelope 2  
FT FT glycoproteins"  
XX  
PN US5663065-A.  
XX  
XX 02-SEP-1997.  
XX  
XX 28-JUN-1991; 91US-0722334.  
XX  
PR 02-JUN-1995; 95US-0459041.  
PR 28-JUN-1991; 91US-0722334.  
PR 19-JUL-1993; 93US-0093453.  
XX  
XX (UYGE-) UNIV GEORGIA STATE RES FOUND.  
XX  
XX Dominguez G, Frey TK, Wang C;  
XX  
XX WPI; 1997-447930/41.  
XX  
XX DNA encoding infectious rubella virus - useful for producing  
PT non-pathogenic mutants for use as vaccines  
XX  
XX Claim 1; Column 9-18; 19pp; English.  
XX  
CC AAT89642 is a cDNA molecule encoding an infectious rubella virus. The  
CC sequence can be modified to give non-pathogenic mutants of this clone.  
CC Such mutants are useful as vaccines for protecting against rubella  
CC infection. The non-pathogenic clones can be used for the vaccination  
CC of pregnant and older women with decreased risk of causing foetal  
CC infection, autoimmune disease or neurological symptoms.  
XX  
SQ Sequence 9759 BP; 1458 A; 3784 C; 3007 G; 1510 T; 0 other;

Query Match 2.8%; Score 42.4; DB 18; Length 9759;  
Best Local Similarity 52.2%; Pred. No. 0.15;  
Matches 94; Conservative 0; Mismatches 86; Indels 0; Gaps 0;  
QY 841 CAGAAAGTCCCGAATTCACACGCTCGCCAAACCGATACCCCGCAACAGCGCG 900  
Db 5600 CATACGGAGCCCGAGTTCGTCGTGGCAGGCCCATTTACACCACCAACCCATCGAG 5659  
QY 901 GGAACCTTACCCAGCAGGAAGGTGCAGAGTTCTTCGGCGCGGCATGTTTCGGCGCGTGG 960

Db 5660 GTCGACTTCACTGATGTTGACATGAACACGACCCCTCGCTACTCTGGAGCGTCCGAGTCGAG 5719  
QY 961 AAGAGCGCGCGCGGATTGACCTTCGGCGCGGACGCGGACCCAGCGCTCGCGCGCGGAC 1020  
Db 5720 ATTAGCGCGCTCTCTTTGGGCTCCCTTCGGCGGAGAACTACCGCGCGCTCGCGCGCGGC 5779  
RESULT 5  
AAA08043  
ID AAA08043 standard; cDNA; 9759 BP.  
XX  
AC AAA08043;  
XX  
DT 19-JUN-2000 (first entry)  
XX  
DE Rubella virus cDNA clone sequence SEQ ID NO:1.  
XX  
KW Rubella virus; infection; infectious; chimeric construct; anti-viral;  
KW vaccine; birth defect; autoimmune disease; ss.  
XX  
OS Rubella virus.  
XX  
PN US6054573-A.  
XX  
XX 25-APR-2000.  
XX  
PF 02-SEP-1997; 97US-0999733.  
XX  
PR 28-JUN-1991; 91US-0722334.  
PR 19-JUL-1993; 93US-0093453.  
PR 02-JUN-1995; 95US-0459041.  
XX  
XX (UYGE-) UNIV GEORGIA STATE.  
XX  
XX Abernathy ES, Pougatchev K, Frey TK;  
XX  
XX WPI; 2000-328366/28.  
XX  
XX Highly infectious rubella virus clones useful for developing a rubella  
PT vaccine that can be safely administered to pregnant and older women  
PT without risk of birth defects and autoimmune disease -  
XX  
XX Disclosure; Column 9-18; 17pp; English.  
XX  
CC The present invention describes nucleic acid molecules (AAA08044 and  
CC AAA08045) which are fragments of the f-Therien rubella virus genome.  
CC AAA08044 and AAA08045 are used to replace the corresponding fragments of  
CC infectious rubella virus cDNA clone with low specific infectivity  
CC referred to as Robo102 to create a chimeric construct with high  
CC specific infectivity, Robo302. The highly infectious rubella virus  
CC clones are useful as molecular biology tools for studying rubella virus  
CC and can be used for developing recombinant vaccines against rubella.  
CC The rubella vaccines developed can be safely administered to pregnant  
CC and older women without risk of birth defects or autoimmune disease.  
CC The present sequence represents a rubella virus cDNA clone used in  
CC the exemplification of the present invention.  
XX  
SQ Sequence 9759 BP; 1458 A; 3784 C; 3007 G; 1510 T; 0 other;

Query Match 2.8%; Score 42.4; DB 21; Length 9759;  
Best Local Similarity 52.2%; Pred. No. 0.15;  
Matches 94; Conservative 0; Mismatches 86; Indels 0; Gaps 0;  
QY 841 CAGAAAGTCCCGAATTCACACGCTCGCCAAACCGATACCCCGCAACAGCGCG 900  
Db 5600 CATACGGAGCCCGAGTTCGTCGTGGCAGGCCCATTTACACCACCAACCCATCGAG 5659  
QY 901 GGAACCTTACCCAGCAGGAAGGTGCAGAGTTCTTCGGCGCGGCATGTTTCGGCGCGTGG 960  
Db 5660 GTCGACTTCACTGATGTTGACATGAACACGACCCCTCGCTACTCTGGAGCGTCCGAGTCGAG 5719  
QY 961 AAGAGCGCGCGCGCATGTTGACCTTCGGCGCGGACGACCCAGCGCTCGCGCGCGGAC 1020

```
Db      57720 ATTAGGCGCCCTCTTGAGGGCCTTCCTTGCACGAAGACTACCGCGCCCTCGCGCGAC 5779
RESULT 6
ID      AAH52083
AAH52083 standard; DNA; 2433 BP.
XX      AC      AAH52083;
XX      DT      04-SEP-2001 (first entry)
DE      Mycobacterium tuberculosis potential drug target gene SEQ ID 137.
KM      Drug target; growth; organism viability; characterisation; ds.
XX      Mycobacterium tuberculosis.
OS      M020013517-A1.
PN      PD      17-MAY-2001.
PP      13-NOV-2000; 2000MO-US31152.
XX      PR      12-NOV-1999; 99US-0165086.
PR      12-NOV-1999; 99US-0165124.
PR      01-FEB-2000; 2000US-0179531.
PA      (REGC ) UNIV CALIFORNIA.
XX      Eisenberg D, Rotstein SH, Marcotte EM;
XX      WPI: 2001-329193/34.
DR      P-PDB; AAG81232.
PT      Identifying nucleotide or polypeptide sequence for use as drug target.
PT      Involves providing algorithm that analyzes a functional relationship
PT      between nucleotide or polypeptide sequences, and comparing the
PT      sequences -
PS      Disclosure; Page 145-146; 207pp; English.
XX      CC      This invention relates to a method for identifying a nucleotide or
XX      CC      polypeptide sequence that may be a drug target, or essential for growth
XX      CC      or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
XX      CC      represent DNA encoding proteins AAG81096 - AAG81241. Mycobacterium
XX      CC      tuberculosis proteins which are potential drug targets. The DNA and
XX      CC      protein sequences are used to illustrate the method of the invention. The
XX      CC      method involves providing an unknown nucleotide or polypeptide sequences,
XX      CC      and comparing it to a number of sequences along with at least one
XX      CC      algorithm capable of analysing a functional relationship between
XX      CC      nucleotide and polypeptide sequences. The method is useful for
XX      CC      characterising the function of nucleic acids and polypeptides that may be
XX      CC      useful as a target for a drug or essential for the growth or viability of
XX      CC      an organism.
SO      Sequence 2433 BP; 468 A; 820 C; 769 G; 376 T; 0 other;
Query Match          2.8%; Score 41.8; DB 22; Length 2433;
Best Local Similarity 48.2%; Pred. No. 0.11;
Matches 118; Conservative 0; Mismatches 127; Indels 0; Gaps 0
QY      835 TTTCAGCAGAAGGCTCCCGAATTCAACGCCGTACGCTGCCAACAGCATACCAGCGCAC 894
Db      1711 TGCCACCAAGGTGTCCTCCGACAGGGCTGGCCAAACCTTGCCCAACGACGATGACAAGAAC 1770
QY      895 AGCGCGGAAAACCTCACCCAGCAGAAAGGTGCAAGATTCTTCGCGCGCGCATGTTCCGC 954
Db      1771 GCCCTGGGACAGCGCACGCGCGCGCGGCTTCGCGCGGCGCGGCGGCTGGATCTGCCGATG 1830
QY      955 CGTTGGAAGAGCGCGCGCGCGCATTTGACCTTCGCGCGCGACGCGCGAGACCCAGCGCTCGGC 1014
Db      1831 TCGGCAAAACCGCACCCAGCAGAGCGCGCACCGGTGCGCGGCTTGAGGCTTACCAACAC 1890
```

Oy	1015	GCACCCGAGAGAAACAATTCGATTACTACGACGCTGACGAGACGAGCCG	1074
Db	1891	CGCTACGCGCGCGGCGCAACTCACTACGACGACTTCGCGGACGATCTGTGTTCC	1950
Oy	1075	TGCCC 1079	
Db	1951	GGCCC 1955	
RESULT 7			
ID	ABL09011	standard; cDNA, 1572 BP.	
XX	ABL09011;		
AC	26-MAR-2002	(first entry)	
XX			
DT			
XX			
DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 21515.		
XX			
KM	Drosophila; developmental biology; cell signalling; insecticide;		
XX	pharmaceutical; gene; se.		
OS	Drosophila melanogaster.		
XX			
PN	WO200171042-A2.		
XX			
PD	27-SEP-2001.		
XX			
PP	23-MAR-2001; 2001WO-US09231.		
XX			
PR	23-MAR-2000; 2000US-191637P.		
XX			
PR	11-JUL-2000; 2000US-0614150.		
XX			
PA	(PEKE ) PE CORP NY.		
XX			
PI	Venter JC, Adams M, Li PWD, Myers EW;		
XX			
DR	WPI: 2001-656860/75.		
XX			
DR	P-PSDB: ABB64908.		
XX			
PT	New isolated nucleic acid detection reagent for detecting 1000 or more		
XX	genes from Drosophila and for elucidating cell signalling and cell-cell		
PT	interactions -		
XX			
BS	Claim 1; SEQ ID NO 21515; 21pp + Sequence Listing; English.		
XX			
CC	The invention relates to an isolated nucleic acid detection reagent		
XX	capable of detecting 1000 or more genes from Drosophila. The invention is		
CC	useful in developmental biology and in elucidating cell signalling and		
XX	cell-cell interactions in higher eukaryotes for the development of		
CC	insecticide, therapeutics and pharmaceutical drugs. The invention		
XX	discloses genomic DNA sequences (AB116176-AB10511), expressed DNA		
CC	sequences (ABL01840-ABL16175) and the encoded proteins		
XX	(AAB57737-ABR72072).		
CC	The sequence data for this patent did not form part of the printed		
XX	specification, but was obtained in electronic format directly from WIPO		
CC	at ftp.wipo.int/pub/published_pct_sequences.		
XX			
SQ	Sequence 1572 BP; 365 A; 465 C; 406 G; 336 T; 0 other;		
Query March 2 8% Score 41.4; DB 23; Length 1572;			
Best local Similarity 47.2% Pred No. 0.12; Indels 141; Gaps 0			
Matches 126; Conservative 0; Mismatches 141; Indels 0; Gaps 0			
Oy	461	CCGCAACCTCGGGGACGACCTTACGCCAATGGGTGGCGGCTTTTACTGTACCAACATCC	520
Db	635	CCGATTAACCTGATGAAGATCTCGACGAGCGCGCTTACCACTTACCAACACCGCCGAGC	694
Oy	521	ATGTGTCCTTTGATTTGATTTGAGGACGACGACGACTTTTGATTCAGTTCACGGATGATA	580
Db	655	GTGAGATCTGCGCCGACATCAAGAGAAAGCTGTGTCTAGTGTGCTCTGGACTTGCAGCAGG	754
Oy	581	TCTCTTGACCTTTGGTCTCCCTCATATCACTCAAGTGCAGGCGCTCAGTGGGTCTGCGGCTC	640

```

Db 755 AGATGCCACCGCTGCCGCTCCACCTCGCTGGAGAAGTCGTACGAGTTGCCGTGACGCC 814
QY 641 CAGGAGATCAGGCTGGTCATGAACACTTCGGGTTCTCTGACACGGCATCTCGACGCCCTCAG 700
Db 815 AGGTGATCACCATTGGCNAAGAGGGCTTCGCTGCCCGGAGGCTCTGTTCCAGCCCTCGT 874
QY 701 TCACAGGCTGGGAGACGACCGTCTTCC 727
Db 875 TCCTGGCATGGAGTCGTGGGCATCC 901

RESULT 8
ABL09010
ID ABL09010 standard; cDNA; 4186 BP.
AC ABL09010;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 21512.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
WPI; 2001-656860/75.
DR P-PSDB; ABB64907.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 21512; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4186 BP; 1092 A; 968 C; 935 G; 1191 T; 0 other;

Query Match 2.8%; Score 41.4; DB 23; Length 4186;
Best Local Similarity 47.2%; Pred. No. 0.19;
Matches 126; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 461 CGCAACCTCGGAGACCTCAGCAATGGGTGGCGCTTTTACTGTACCACTCC 520
Db 2188 CGATTACTCTGATGAAGATCTTGACGAGCGCGGTACAGCTCACCAACCGCGGAGC 2247
QY 521 ATGGGTGCTTTCTGATTTGGTAGCAGCAGCACTTCTTGGATCAGTTCACGGATGATA 580

```

```

Db 2248 GTGAGATCGTCCGCGCATCAGAGAGCTGTGTACTCGTGGCTCTGGACTTCGAGCAGG 2307
QY 581 TCTCTTCGACCTTTTGGTTCTCTCCATCACTCAGGTGCAGCGCTCAGTGGGTCTGCGCGTC 640
Db 2308 AGATGGCCACCGCTGCCGCTCCACCTCGCTGGAGAAGTCGTACGAGTTGCTGACGGCC 2367
QY 641 CAGGATCAGGCTGGTCATGAACACTTCGGGTTCTCTGACCGCATCTCGCAGGCCCTCAG 700
Db 2368 AGGTGATCACCATTGGCAACGAGCGCTTCGCTGCCCGGAGGCTCTGTTCAGGCCCTCGT 2427
QY 701 TCACAGGCTGGGAGACGACCGTCTTCC 727
Db 2428 TCCTGGGCATGGAGTCGTGGGCATCC 2454

RESULT 9
ABK95664
ID ABK95664 standard; cDNA; 840 BP.
XX
AC ABK95664;
XX
DT 24-SEP-2002 (first entry)
XX
DE Grass cDNA for allergen p 5 mutant #13.
XX
KW Immunoglobulin E; IgE; allergen; allergy; ss; gene; hay fever;
KW rhinoconductivitis; rhinitis; asthma; systemic anaphylaxis; mutant;
KW vaccine; antiallergic; B cell epitope.
XX
OS Phleum sp.
OS Synthetic.
XX
PN WO200240676-A2.
XX
PD 23-MAY-2002.
XX
PF 16-NOV-2001; 2001WO-DK00764.
XX
PR 16-NOV-2000; 2000DK-0001718.
PR 16-NOV-2000; 2000US-249361P.
PR 14-JUN-2001; 2001US-298170P.
XX
PA (ALKA-) ALK-ABELLO AS.
XX
PI Holm J, Ipeen H, Nedergaard Larsen J, Spangfort MD;
XX
WPI; 2002-508328/54.
DR P-PSDB; ABG67048.
XX
PT New recombinant mutant allergen, useful for preventing and/or treating
PT allergy, comprises multiple mutations and reduced immunoglobulin E
PT binding affinity -
XX
PS Example 8; Page 119-120; 210pp; English.
XX
CC The invention relates to a recombinant allergen (I) which is a mutant of
CC a naturally occurring allergen, where the mutant allergen has at least
CC four primary mutations, which each reduce the specific immunoglobulin E
CC (IgE) binding capability of the mutated allergen as compared to the IgE
CC binding capability of the naturally occurring allergen, where each
CC primary mutation is a substitution of one surface-exposed amino acid
CC residue with another residue, which does not occur in the same position
CC in the amino acid sequence of any known homologous protein within the
CC taxonomic species from which the naturally occurring allergen
CC originates, and each primary mutation is spaced from each other primary
CC mutation by at least 15 Angstrom , and the primary mutations are placed
CC in such a manner that at least one circular surface region with a area
CC of 800 Angstrom ^2 comprises no mutation. Also included are a composition
CC comprising two or more of the recombinant allergens, where the variant
CC allergen is defined by having at least one primary mutation, which is
CC absent in at least one of the other variants, and for each variant no
CC secondary mutation is present within a radius of 15 Angstrom from each
CC absent primary mutation; a DNA sequence encoding the recombinant allergen

```



or its derivative, partial sequence or degenerated sequence, or a sequence which hybridizes to it under stringent conditions, where the derivative, partial sequence, degenerated sequence or hybridising sequence encodes a peptide having at least one B cell epitope; an expression vector comprising the DNA and a host cell comprising the vector. The recombinant allergen is useful as a pharmaceutical, for preparing a pharmaceutical for preventing and/or treating allergy, or in a diagnostic assay for assessing relevance, safety or outcome of therapy of a subject, where an IGE containing sample of the subject is mixed with the recombinant allergen and assessed for the level of reactivity between the IGE in the sample and the recombinant allergen. The recombinant allergen or compositions are useful for generating an immune response in a subject, for vaccination or treatment of a subject or for the treatment, prevention or alleviation of allergic reactions in a subject e.g. hay fever, rhinoconjunctivitis, rhinitis, asthma or systemic anaphylaxis. The present sequence is a cDNA encoding a recombinant allergen of the invention.

Sequence 840 BP; 164 A; 340 C; 249 G; 87 T; 0 other;

Query Match 2.8%; Score 41.2; DB 24; Length 840;

Best Local Similarity 48.3%; Pred. No. 0.1;

Matches 115; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

```
OY 847 GTCGCCGAATTCAACGCTGACGCTGCGCAACGCGGATACCGCGGAACGCGGGAAAC 906
DB 385 GCCACCGTAGAGAGCGCTCTCGCATCTGCGCGGACCTTCAGGTCCACGCGGTCAAG 444
OY 907 CTCACCCAGCAGAGAGTGACAGATTCTCTGCGCGCGGCGCATGTTGCGCGCTTGAAGAC 966
DB 445 CCGCGGCGCGAGAGAGTCAAGTCAATCCCGCGCGAGCTGCAGGTCAATCGAAGAGTTC 504
OY 967 GGGCGCGCGATGACCTCGCGCGCGGAGACGACCGGATCGCGCGCGGCGGCGGCGGAG 1026
DB 505 GACGCGCGCTTAAAGTGTGCTGCGACGCGCGCGCAACGCGCGCGCGCAACGCAAGTTC 564
OY 1027 AGGAACAACAATTGATTACTACAGACACGCTGACGAGACGAGCGGCTGCGCGCTTCG 1084
DB 565 ACCGTTCTGAGCGCGCTTCAACGAGCGCATCAAGGCGAGGAGCGGCGGCGGCTTACG 622
```

RESULT 10

ABK95652 standard; cDNA; 861 BP.

ABK95652;

24-SEP-2002 (first entry)

Grass CDNA for allergen p 5 mutant #1.

Immunoglobulin E; IGE; allergen; allergy; ss; gene; hay fever;  
rhinoconjunctivitis; rhinitis; asthma; systemic anaphylaxis; mutant;  
vaccine; anti-allergic; B cell epitope.

Phleum sp.

Synthetic.

WO200240676-A2.

23-MAY-2002.

16-NOV-2001; 2001WO-DK00764.

16-NOV-2000; 2000DK-0001718.

16-NOV-2000; 2000US-249361P.

14-JUN-2001; 2001US-298170P.

(ALK-A) ALK-ABELLO AS.

Holm J, Ipsen H, Nedergaard Larsen J, Spangfort MD;

WPI; 2002-508328/54.

DR P-PSDB; ABG67036.

XX

PT New recombinant mutant allergen, useful for preventing and/or treating

PT allergy, comprises multiple mutations and reduced immunoglobulin E

PT binding affinity -

PS Example 8; Page 115; 210p; English.

The invention relates to a recombinant allergen (I) which is a mutant of a naturally occurring allergen, where the mutant allergen has at least four primary mutations, which each reduce the specific immunoglobulin E (IGE) binding capability of the mutated allergen as compared to the IGE binding capability of the naturally occurring allergen, where each CC primary mutation is a substitution of one surface-exposed amino acid residue with another residue, which does not occur in the same position in the amino acid sequence of any known homologous protein within the CC taxonomic species from which the naturally occurring allergen CC originates, and each primary mutation is spaced from each other primary mutation by at least 15 Angstrom, and the primary mutations are placed in such a manner that at least one circular surface region with a area of 800 Angstrom<sup>2</sup> comprises no mutation. Also included are a composition comprising two or more of the recombinant allergens, where the variant allergen is defined by having at least one primary mutation, which is absent in at least one of the other variants, and for each variant no secondary mutation is present within a radius of 15 Angstrom from each CC absent primary mutation; a DNA sequence encoding the recombinant allergen or its derivative, partial sequence or degenerated sequence, or a sequence which hybridizes to it under stringent conditions, where the CC derivative, partial sequence, degenerated sequence or hybridising CC sequence encodes a peptide having at least one B cell epitope; an CC expression vector comprising the DNA and a host cell comprising the CC vector. The recombinant allergen is useful as a pharmaceutical for preparing a pharmaceutical for preventing and/or treating allergy, or in a diagnostic assay for assessing relevance, safety or outcome of therapy of a subject, where an IGE containing sample of the subject is mixed with the recombinant allergen and assessed for the level of reactivity between the IGE in the sample and the recombinant allergen. The CC recombinant allergen or compositions are useful for generating an CC immune response in a subject, for vaccination or treatment of a subject CC or for the treatment, prevention or alleviation of allergic reactions CC in a subject e.g. hay fever, rhinoconjunctivitis, rhinitis, asthma or CC systemic anaphylaxis. The present sequence is a cDNA encoding a CC recombinant allergen of the invention.

Sequence 861 BP; 172 A; 346 C; 251 G; 92 T; 0 other;

Query Match 2.8%; Score 41.2; DB 24; Length 861;

Best Local Similarity 48.3%; Pred. No. 0.1; Matches 115; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

```
OY 847 GTCGCCGAATTCAACGCTGACGCTGCGCAACGCGGATACCGCGGAACGCGGGAAAC 906
DB 385 GCCACCGTAGAGAGCGCTCTCGCATCTGCGCGGACCTTCAGGTCCACGCGGTCAAG 444
OY 907 CTCACCCAGCAGAGAGTGACAGATTCTCTGCGCGCGGCGCATGTTGCGCGCTTGAAGAC 966
DB 445 CCGCGGCGCGAGAGAGTCAAGTCAATCCCGCGCGAGCTGCAGGTCAATCGAAGAGTTC 504
OY 967 GGGCGCGCGATGACCTCGCGCGCGGAGACGACCGGATCGCGCGCGGCGGCGGCGGAG 1026
DB 505 GACGCGCGCTTAAAGTGTGCTGCGACGCGCGCGCAACGCGCGCGCGCAACGCAAGTTC 564
OY 1027 AGGAACAACAATTGATTACTACAGACACGCTGACGAGACGAGCGGCTGCGCGCTTCG 1084
DB 565 ACCGTTCTGAGCGCGCTTCAACGAGCGCATCAAGGCGAGGAGCGGCGGCGGCTTACG 622
```

RESULT 11

ABK95653 standard; cDNA; 861 BP.

ABK95653;

XX

DT 24-SEP-2002 (first entry)  
 XX Grass cDNA for allergen p 5 mutant #2.  
 XX Immunoglobulin E; IgE; allergen; allergy; ss; gene; hay fever;  
 KW rhinoconjunctivitis; rhinitis; asthma; systemic anaphylaxis; mutant;  
 XX vaccine; anti-allergic; B cell epitope.  
 XX Phleum sp.  
 OS Synthetic.  
 XX WO200240676-A2.  
 PN 23-MAY-2002.  
 XX 16-NOV-2001; 2001WO-DK00764.  
 XX 16-NOV-2000; 2000DK-0001718.  
 PR 16-NOV-2000; 2000US-249361P.  
 PR 14-JUN-2001; 2001US-298170P.  
 XX (ALKA-) ALK-ABELLO AS.  
 PA Holm J, Ipsen H, Nedergaard Larsen J, Spangfort MD;  
 PI WPI; 2002-508328/54.  
 XX P-PSDB; ABG67037.  
 DR New recombinant mutant allergen, useful for preventing and/or treating  
 PT allergy, comprises multiple mutations and reduced immunoglobulin E  
 PT binding affinity -  
 XX Example 8; Page 115-116; 210pp; English.  
 XX The invention relates to a recombinant allergen (I) which is a mutant of  
 CC a naturally occurring allergen, where the mutant allergen has at least  
 CC four primary mutations, which each reduce the specific immunoglobulin E  
 CC (IgE) binding capability of the mutated allergen as compared to the IgE  
 CC binding capability of the naturally occurring allergen, where each  
 CC primary mutation is a substitution of one surface-exposed amino acid  
 CC residue with another residue, which does not occur in the same position  
 CC in the amino acid sequence of any known homologous protein within the  
 CC taxonomic species from which the naturally occurring allergen  
 CC originates, and each primary mutation is spaced from each other primary  
 CC mutation by at least 15 Angstrom, and the primary mutations are placed  
 CC in such a manner that at least one circular surface region with a area  
 CC of 800 Angstrom<sup>2</sup> comprises no mutation. Also included are a composition  
 CC comprising two or more of the recombinant allergens, where the variant  
 CC allergen is defined by having at least one primary mutation, which is  
 CC absent in at least one of the other variants, and for each variant no  
 CC secondary mutation is present within a radius of 15 Angstrom from each  
 CC absent primary mutation; a DNA sequence encoding the recombinant allergen  
 CC or its derivative, partial sequence or degenerated sequence, or a  
 CC sequence which hybridises to it under stringent conditions, where the  
 CC derivative, partial sequence, degenerated sequence or hybridising  
 CC sequence encodes a peptide having at least one B cell epitope; an  
 CC expression vector comprising the DNA and a host cell comprising the  
 CC vector. The recombinant allergen is useful as a pharmaceutical, for  
 CC preparing a pharmaceutical for preventing and/or treating allergy, or in  
 CC a diagnostic assay for assessing relevance, safety or outcome of therapy  
 CC of a subject, where an IgE containing sample of the subject is mixed  
 CC with the recombinant allergen and assessed for the level of reactivity  
 CC between the IgE in the sample and the recombinant allergen. The  
 CC recombinant allergen or compositions are useful for generating an  
 CC immune response in a subject, for vaccination or treatment of a subject  
 CC or for the treatment, prevention or alleviation of allergic reactions  
 CC in a subject e.g. hay fever, rhinoconjunctivitis, rhinitis, asthma or  
 CC systemic anaphylaxis. The present sequence is a cDNA encoding a  
 CC recombinant allergen of the invention.  
 XX Sequence 861 BP; 171 A; 348 C; 249 G; 93 T; 0 other;  
 SQ Query Match 2.8%; Score 41.2; DB 24; Length 861;

Best Local Similarity 48.3%; Pred. No. 0.1;  
 Matches 115; Conservative 0; Mismatches 123; Indels 0; Gaps 0;  
 QY 847 GTCCCCGGAATTCAACGCGGTACACGCTCGCCCAACGCGATACCCGCGAAGCAGCGCGGAAAC 906  
 DB 385 GCCACCGTAAGCAGCGCGCTCCGCATCATCGCGGCACCCCTCGAGGTCCACGCCGTCAAG 444  
 QY 907 TTCACCCAGCAGGAAGGTGCAGATTCTCTCGCGCGCGCATGTTCCGCCGTGGAAGAGC 966  
 DB 445 CCCGCGCGGAGGAGGTCAAGGTTCATCCCGCGCGAGCTGCAGGTTCATCGAAGGTC 504  
 QY 967 GCGCGCGGATTGACCTCGCGCGAGCGGAGCACCCAGCGCTCGCGCGACCCGCGAG 1026  
 DB 505 GACGCGCGCTTCAAGGTTCGCTGCCACCGCGCCNAGCGCGCCCGCCCAACGACGATT 564  
 QY 1027 AGGAACAAACAATTTTCGATTACTCAGACACGCTGACGCGAGCAGCGCGTGCCTTCG 1084  
 DB 565 ACCGTCTTCGAGCGCGCTTCAACGACGCGCATCAAGGCGAGCAGCGCGCGCGCTTACG 622

RESULT 12  
 ABK95654  
 ID ABK95654 standard; cDNA; 861 BP.  
 XX AC ABK95654;  
 XX DT 24-SEP-2002 (first entry)  
 XX DE Grass cDNA for allergen p 5 mutant #3.  
 XX KW Immunoglobulin E; IgE; allergen; allergy; ss; gene; hay fever;  
 KW rhinoconjunctivitis; rhinitis; asthma; systemic anaphylaxis; mutant;  
 KW vaccine; anti-allergic; B cell epitope.  
 XX OS Phleum sp.  
 OS Synthetic.  
 XX PN WO200240676-A2.  
 XX PD 23-MAY-2002.  
 XX 16-NOV-2001; 2001WO-DK00764.  
 PR 16-NOV-2000; 2000DK-0001718.  
 PR 16-NOV-2000; 2000US-249361P.  
 PR 14-JUN-2001; 2001US-298170P.  
 XX (ALKA-) ALK-ABELLO AS.  
 XX Holm J, Ipsen H, Nedergaard Larsen J, Spangfort MD;  
 PI WPI; 2002-508328/54.  
 XX P-PSDB; ABG67038.  
 XX New recombinant mutant allergen, useful for preventing and/or treating  
 PT allergy, comprises multiple mutations and reduced immunoglobulin E  
 PT binding affinity -  
 XX Example 8; Page 116; 210pp; English.  
 XX The invention relates to a recombinant allergen (I) which is a mutant of  
 CC a naturally occurring allergen, where the mutant allergen has at least  
 CC four primary mutations, which each reduce the specific immunoglobulin E  
 CC (IgE) binding capability of the mutated allergen as compared to the IgE  
 CC binding capability of the naturally occurring allergen, where each  
 CC primary mutation is a substitution of one surface-exposed amino acid  
 CC residue with another residue, which does not occur in the same position  
 CC in the amino acid sequence of any known homologous protein within the  
 CC taxonomic species from which the naturally occurring allergen  
 CC originates, and each primary mutation is spaced from each other primary  
 CC mutation by at least 15 Angstrom, and the primary mutations are placed  
 CC in such a manner that at least one circular surface region with a area  
 CC of 800 Angstrom<sup>2</sup> comprises no mutation. Also included are a composition  
 CC comprising two or more of the recombinant allergens, where the variant  
 CC allergen is defined by having at least one primary mutation, which is  
 CC absent in at least one of the other variants, and for each variant no  
 CC secondary mutation is present within a radius of 15 Angstrom from each  
 CC absent primary mutation; a DNA sequence encoding the recombinant allergen  
 CC or its derivative, partial sequence or degenerated sequence, or a  
 CC sequence which hybridises to it under stringent conditions, where the  
 CC derivative, partial sequence, degenerated sequence or hybridising  
 CC sequence encodes a peptide having at least one B cell epitope; an  
 CC expression vector comprising the DNA and a host cell comprising the  
 CC vector. The recombinant allergen is useful as a pharmaceutical, for  
 CC preparing a pharmaceutical for preventing and/or treating allergy, or in  
 CC a diagnostic assay for assessing relevance, safety or outcome of therapy  
 CC of a subject, where an IgE containing sample of the subject is mixed  
 CC with the recombinant allergen and assessed for the level of reactivity  
 CC between the IgE in the sample and the recombinant allergen. The  
 CC recombinant allergen or compositions are useful for generating an  
 CC immune response in a subject, for vaccination or treatment of a subject  
 CC or for the treatment, prevention or alleviation of allergic reactions  
 CC in a subject e.g. hay fever, rhinoconjunctivitis, rhinitis, asthma or  
 CC systemic anaphylaxis. The present sequence is a cDNA encoding a  
 CC recombinant allergen of the invention.  
 XX Sequence 861 BP; 171 A; 348 C; 249 G; 93 T; 0 other;  
 SQ Query Match 2.8%; Score 41.2; DB 24; Length 861;

comprising two or more of the recombinant allergens, where the variant allergen is defined by having at least one primary mutation, which is absent in at least one of the other variants, and for each variant no secondary mutation is present within a radius of 15 Angstrom from each absent primary mutation; a DNA sequence encoding the recombinant allergen or its derivative, partial sequence or degenerated sequence, where the sequence which hybridizes to it under stringent conditions, where the derivative, partial sequence, degenerated sequence or hybridizing sequence encodes a peptide having at least one B cell epitope; an expression vector comprising the DNA and a host cell comprising the vector. The recombinant allergen is useful as a pharmaceutical, for preparing a pharmaceutical for preventing and/or treating allergy, or in a diagnostic assay for assessing relevance, safety or outcome of therapy of a subject, where an IGE containing sample of the subject is mixed with the recombinant allergen and assessed for the level of reactivity between the IGE in the sample and the recombinant allergen. The recombinant allergen or compositions are useful for generating an immune response in a subject, for vaccination or treatment of a subject or for the treatment, prevention or alleviation of allergic reactions in a subject e.g. hay fever, rhinoconjunctivitis, rhinitis, asthma or systemic anaphylaxis. The present sequence is a cDNA encoding a recombinant allergen of the invention.

Sequence 861 BP; 173 A; 344 C; 252 G; 92 T; 0 other;

Query Match 2.8%; Score 41.2; DB 24; Length 861;

Best Local Similarity 48.3%; Pred. No. 0.1; Mismatches 123; Indels 0; Gaps 0;

Matches 115; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

847 GTCCCGAATTCAACGCGTACACGCTCCCAACCGATCCCGGAACACCGCGGAAC 906  
385 GCCACCGTAAAGGAGCGCTCAAGTCAATCCCGGCAACCTCCAGGTCACCGCTCAAG 444  
907 CTCACCCAGCAGAGAGTGCAGAGTTCTCGCGCGCGGCATGTTCCGCCCTTGAAGACC 966  
445 CCGCGCGCGGAGAGAGTCAAGTCAATCCCGGCAACCGATCCCGGAACACCGCGGAAC 1026  
967 GCGCGCGCGGATTCAGCTCGCGCGGAGAGACCGATCCCGGCGCGCGCGGAG 1026  
505 GACGCGCGCTTCAAGGTGCTGCGCCGCGGCAACCGCGCGCGCGGCAACGAGATT 564  
1027 AGGAAACAATTTGATTACTCAGACACCGTGAACGAGAGAGCGGCTGCGCTTCG 1084  
565 ACCGCTTGAAGCGCGCTTCAACGAGCCCATCAAGGCGAGAGCGCGCGCGCTACG 622

RESULT 13

ABK95657  
ID ABK95657 standard; cDNA; 861 BP.

XX ABK95657;

XX 24-SBP-2002 (first entry)

XX Grass cDNA for allergen p 5 mutant #6.

XX Immunoglobulin E; IGE; allergen; allergy; 86; gene; hay fever;  
XX rhinoconjunctivitis; rhinitis; asthma; systemic anaphylaxis; mutant;  
XX vaccine; anti-allergic; B cell epitope.

XX Phleum sp.

XX OS Synthetic.

XX PN WO200240676-A2.

XX PD 23-MAY-2002.

XX PF 16-NOV-2001; 2001WO-DK00764.

XX PR 16-NOV-2000; 2000DK-0001718.

XX PR 16-NOV-2000; 2000US-249361P.

XX PR 14-JUN-2001; 2001US-298170P.

XX XX

PA (ALKA-) ALK-ABELLO AS.

XX Holm J, Ipsen H, Nedergaard Larsen J, Spangfort MD;

XX WPI; 2002-508328/54.

XX P-PSDB; AB667041.

PT New recombinant mutant allergen, useful for preventing and/or treating  
PT allergy, comprises multiple mutations and reduced immunoglobulin E  
PT binding affinity -

XX Example 8; Page 117; 210pp; English.

XX The invention relates to a recombinant allergen (II) which is a mutant of  
XX a naturally occurring allergen, where the mutant allergen has at least  
XX four primary mutations, which each reduce the specific immunoglobulin E  
XX (IGE) binding capability of the mutated allergen as compared to the IGE  
XX binding capability of the naturally occurring allergen, where each  
XX primary mutation is a substitution of one surface-exposed amino acid  
XX residue with another residue, which does not occur in the same position  
XX in the amino acid sequence of any known homologous protein within the  
XX taxonomic species from which the naturally occurring allergen  
XX originates, and each primary mutation is spaced from each other primary  
XX mutation by at least 15 Angstrom, and the primary mutations are placed  
XX in such a manner that at least one circular surface region with a area  
XX of 800 Angstrom<sup>2</sup> comprises no mutation. Also included are a composition  
XX comprising two or more of the recombinant allergens, where the variant  
XX allergen is defined by having at least one primary mutation, which is  
XX absent in at least one of the other variants, and for each variant no  
XX secondary mutation is present within a radius of 15 Angstrom from each  
XX absent primary mutation; a DNA sequence encoding the recombinant allergen  
XX or its derivative, partial sequence or degenerated sequence, or a  
XX sequence which hybridizes to it under stringent conditions, where the  
XX derivative, partial sequence, degenerated sequence or hybridizing  
XX sequence encodes a peptide having at least one B cell epitope; an  
XX expression vector comprising the DNA and a host cell comprising the  
XX vector. The recombinant allergen is useful as a pharmaceutical, for  
XX preparing a pharmaceutical for preventing and/or treating allergy, or in  
XX a diagnostic assay for assessing relevance, safety or outcome of therapy  
XX of a subject, where an IGE containing sample of the subject is mixed  
XX with the recombinant allergen and assessed for the level of reactivity  
XX between the IGE in the sample and the recombinant allergen. The  
XX recombinant allergen or compositions are useful for generating an  
XX immune response in a subject, for vaccination or treatment of a subject  
XX or for the treatment, prevention or alleviation of allergic reactions  
XX in a subject e.g. hay fever, rhinoconjunctivitis, rhinitis, asthma or  
XX systemic anaphylaxis. The present sequence is a cDNA encoding a  
XX recombinant allergen of the invention.

SQ Sequence 861 BP; 171 A; 346 C; 252 G; 92 T; 0 other;

Query Match 2.8%; Score 41.2; DB 24; Length 861;

Best Local Similarity 48.3%; Pred. No. 0.1; Mismatches 123; Indels 0; Gaps 0;

Matches 115; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

847 GTCCCGAATTCAACGCGTACACGCTCCCAACCGATCCCGGAACACCGCGGAAC 906  
385 GCCACCGTAAAGGAGCGCTCAAGTCAATCCCGGCAACCTCCAGGTCACCGCTCAAG 444  
907 CTCACCCAGCAGAGAGTGCAGAGTTCTCGCGCGCGGCATGTTCCGCCCTTGAAGACC 966  
445 CCGCGCGCGGAGAGAGTCAAGTCAATCCCGGCAACCGATCCCGGAACACCGCGGAAC 1026  
967 GCGCGCGCGGATTCAGCTCGCGCGGAGAGACCGATCCCGGCGCGCGCGGAG 1026  
505 GACGCGCGCTTCAAGGTGCTGCGCCGCGGCAACCGCGCGCGCGGCAACGAGATT 564  
1027 AGGAAACAATTTGATTACTCAGACACCGTGAACGAGAGAGCGGCTGCGCTTCG 1084  
565 ACCGCTTGAAGCGCGCTTCAACGAGCCCATCAAGGCGAGAGCGCGCGCGCTACG 622

RESULT 14

ABK95658  
ID ABK95658 standard; cDNA; 861 BP.  
XX  
AC ABK95658;  
XX  
DT 24-SEP-2002 (first entry)  
XX  
DE Grass cDNA for allergen p 5 mutant #7.  
XX  
KW Immunoglobulin E; IgE; allergen; allergy; ss; gene; hay fever;  
KW rhinoconductivitis; rhinitis; asthma; systemic anaphylaxis; mutant;  
KW vaccine; antiallergic; B cell epitope.  
XX  
OS Phleum sp.  
OS Synthetic.  
XX  
XX WO200240676-A2.  
XX  
XX 23-MAY-2002.  
XX  
XX 16-NOV-2001; 2001WO-DK00764.  
XX  
XX 16-NOV-2000; 2000DX-0001718.  
PR 16-NOV-2000; 2000US-249361P.  
PR 14-JUN-2001; 2001US-298170P.  
XX  
XX (ALKA-) ALK-ABELLO AS.  
XX  
XX Holm J, Ipsen H, Nedergaard Larsen J, Spangfort MD;  
XX  
XX WPI; 2002-508328/54.  
XX  
XX P-PSDB; ABG67042.  
XX  
XX New recombinant mutant allergen, useful for preventing and/or treating  
PT allergy, comprises multiple mutations and reduced immunoglobulin E  
PT binding affinity -  
XX  
XX Example 8; Page 117; 210pp; English.  
XX  
XX The invention relates to a recombinant allergen (I) which is a mutant of  
CC a naturally occurring allergen, where the mutant allergen has at least  
CC four primary mutations, which each reduce the specific immunoglobulin E  
CC (IGE) binding capability of the mutated allergen as compared to the IGE  
CC binding capability of the naturally occurring allergen, where each  
CC primary mutation is a substitution of one surface-exposed amino acid  
CC residue with another residue, which does not occur in the same position  
CC in the amino acid sequence of any known homologous protein within the  
CC taxonomic species from which the naturally occurring allergen  
CC originates, and each primary mutation is spaced from each other primary  
CC mutation by at least 15 Angstrom, and the primary mutations are placed  
CC in such a manner that at least one circular surface region with a area  
CC of 800 Angstrom<sup>2</sup> comprises no mutation. Also included are a composition  
CC comprising two or more of the recombinant allergens, where the variant  
CC allergen is defined by having at least one primary mutation, which is  
CC absent in at least one of the other variants, and for each variant no  
CC secondary mutation is present within a radius of 15 Angstrom from each  
CC absent primary mutation; a DNA sequence encoding the recombinant allergen  
CC or its derivative, partial sequence or degenerated sequence, or a  
CC sequence which hybridizes to it under stringent conditions, where the  
CC derivative, partial sequence, degenerated sequence or hybridising  
CC sequence encodes a peptide having at least one B cell epitope; an  
CC expression vector comprising the DNA and a host cell comprising the  
CC vector. The recombinant allergen is useful as a pharmaceutical, for  
CC preparing a pharmaceutical for preventing and/or treating allergy, or in  
CC a diagnostic assay for assessing relevance, safety or outcome of therapy  
CC of a subject, where an IGE containing sample of the subject is mixed  
CC with the recombinant allergen and assessed for the level of reactivity  
CC between the IGE in the sample and the recombinant allergen. The  
CC recombinant allergen or compositions are useful for generating an  
CC immune response in a subject, for vaccination or treatment of a subject  
CC or for the treatment, prevention or alleviation of allergic reactions  
CC in a subject e.g. hay fever, rhinoconductivitis, rhinitis, asthma or  
CC systemic anaphylaxis. The present sequence is a cDNA encoding a

CC recombinant allergen of the invention.  
XX  
SQ Sequence 861 BP; 170 A; 345 C; 254 G; 92 T; 0 other;  
XX  
Query Match 2.8%; Score 41.2; DB 24; Length 861;  
XX Best Local Similarity 48.3%; Pred. No. 0.1; Indels 0; Gaps 0;  
XX Matches 115; Conservative 0; Mismatches 123;  
QY 847 GTCCCGAATTCACGCGTACACGCTCGCAACGCGATACCCGCAACAGCGCGGAAAC 906  
DB 385 GCCACCGTAAAGCAGCGGCTCCGCATCATCGCGGCACCTCGAGGTCCAGCGGTCAAG 444  
QY 907 CTCACCCAGCAGGAGGTGCAGAGTTCTCTCGGCGCGCATGTTCTGCGCGTTCGAAGAGC 966  
DB 445 CCGCGGCGGAGGAGGTCAAGGTTCATCCCGCGGAGCTCCAGGTATCGAGAGAGTC 504  
QY 967 GCGCGCGCGATTGACCTCGCGCGGCGGACGACCCAGCGCTCGGCGCGACCCGCGAG 1026  
DB 505 GACGCGCGCTTCAAGGTCTGTCACCGCGCGCAACGCGCGCCCGCAACGCAAGATT 564  
QY 1027 AGGACACAAATTTTCGATTACTCAGACACGCTCAGCGAGCAGCGCTGCCCTTCG 1084  
DB 565 ACCGTCTTCGAGGCGCGCTTCAACGACGACGCATCAAGCGAGCAGCGCGCGCTTACG 622  
XX  
RESULT 15  
ABK95659  
ID ABK95659 standard; cDNA; 861 BP.  
XX  
AC ABK95659;  
XX  
XX 24-SEP-2002 (first entry)  
XX  
XX Grass cDNA for allergen p 5 mutant #8.  
KW Immunoglobulin E; IgE; allergen; allergy; ss; gene; hay fever;  
KW rhinoconductivitis; rhinitis; asthma; systemic anaphylaxis; mutant;  
KW vaccine; antiallergic; B cell epitope.  
XX  
XX Phleum sp.  
OS Synthetic.  
XX  
XX WO200240676-A2.  
XX  
XX 23-MAY-2002.  
XX  
XX 16-NOV-2001; 2001WO-DK00764.  
XX  
XX 16-NOV-2000; 2000DX-0001718.  
PR 16-NOV-2000; 2000US-249361P.  
PR 14-JUN-2001; 2001US-298170P.  
XX  
XX (ALKA-) ALK-ABELLO AS.  
XX  
XX Holm J, Ipsen H, Nedergaard Larsen J, Spangfort MD;  
XX  
XX WPI; 2002-508328/54.  
XX  
XX P-PSDB; ABG67043.  
XX  
XX New recombinant mutant allergen, useful for preventing and/or treating  
PT allergy, comprises multiple mutations and reduced immunoglobulin E  
PT binding affinity -  
XX  
XX Example 8; Page 118; 210pp; English.  
XX  
XX The invention relates to a recombinant allergen (I) which is a mutant of  
CC a naturally occurring allergen, where the mutant allergen has at least  
CC four primary mutations, which each reduce the specific immunoglobulin E  
CC (IGE) binding capability of the mutated allergen as compared to the IGE  
CC binding capability of the naturally occurring allergen, where each  
CC primary mutation is a substitution of one surface-exposed amino acid  
CC residue with another residue, which does not occur in the same position  
CC in the amino acid sequence of any known homologous protein within the

CC taxonomic species from which the naturally occurring allergen  
CC originates, and each primary mutation is spaced from each other primary  
CC mutation by at least 15 Angstrom, and the primary mutations are placed  
CC in such a manner that at least one circular surface region with a area  
CC of 800 Angstrom<sup>2</sup> comprises no mutation. Also included are a composition  
CC comprising two or more of the recombinant allergens, where the variant  
CC allergen is defined by having at least one primary mutation, which is  
CC absent in at least one of the other variants, and for each variant no  
CC secondary mutation is present within a radius of 15 Angstrom from each  
CC absent primary mutation; a DNA sequence encoding the recombinant allergen  
CC or its derivative, partial sequence or degenerated sequence, or a  
CC sequence which hybridises to it under stringent conditions, where the  
CC derivative, partial sequence, degenerated sequence or hybridising  
CC sequence encodes a peptide having at least one B cell epitope; an  
CC expression vector comprising the DNA and a host cell comprising the  
CC vector. The recombinant allergen is useful as a pharmaceutical, for  
CC preparing a pharmaceutical for preventing and/or treating allergy, or in  
CC a diagnostic assay for assessing relevance, safety or outcome of therapy  
CC of a subject, where an 198 containing sample of the subject is mixed  
CC with the recombinant allergen and assessed for the level of reactivity  
CC between the 198 in the sample and the recombinant allergen. The  
CC recombinant allergen or compositions are useful for generating an  
CC immune response in a subject, for vaccination or treatment of a subject  
CC or for the treatment, prevention or alleviation of allergic reactions  
CC in a subject e.g. hay fever, rhinoconjunctivitis, rhinitis, asthma or  
CC systemic anaphylaxis. The present sequence is a cDNA encoding a  
CC recombinant allergen of the invention.

XX  
SQ Sequence 861 BP; 170 A; 348 C; 251 G; 92 T; 0 other;

Query Match 2.8%; Score 41.2; DB 24; Length 861;

Best Local Similarity 48.3%; Pred. No. 0.1;

Matches 115; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 847 GTCCCGCAATTCAAGCGCTACAGCTCGCCCAAGCGGATCCCGGAAACAGCGCGGAAAC 906  
DB 385 GCCACCGTAGAGAGCGCGCTCGCATCATCGCGCAACCCCTCGAGGTCCACGCGCTCAAG 444  
QY 907 CTCACCCAGAGAGAGTGCAGATTCTCGGCGCGGCATGTTCCGCGCTTGAAGAGC 966  
DB 445 CCGCGCGCGAGAGAGTCAAGTCTATCCCCCGGCGAGCTGACGTATCGAAGAGTC 504  
QY 967 GCGCGCGCGATTGACTCGCGCGGAGCGGAGCAAGCGCTCGGCGCGCAACCCGACG 1026  
DB 505 GAGCGCGCTTCAAGGTCTGTCACCGCGCCCAAGCGCGCGCCCGCAACGAAAGATT 564  
QY 1027 AGGAACAACATTTGATTAATCAGACAGCGTGAAGCAAGCGCGCTGCGCTTGG 1084  
DB 565 ACCGCTTTCAGAGCGCGCTTCAAGACCGCATCAAGGAGAGAGCGGCGCGCTACG 622

Search completed: December 4, 2002, 21:53:19  
Job time : 334 secs

**THIS PAGE BLANK (USPTO)**



High quality sequence stop: 543.

FEATURES  
source  
1. 785  
/organism="Hordeum vulgare"  
/cultivar="Morex"  
/db\_xref="taxon:4513"  
/clone="HVSMB10003M12f"  
/HVCNAD010 (20 DAP)"  
/tissue\_type="20 DAP spike"  
/lab\_host="SOLR"  
/note="vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;  
Plants were grown in the greenhouse at the University of  
California, Riverside (Penton, SJ Close, TJ Close). Whole  
spikes with awns trimmed were collected at 20 DAP (Penton  
). Total RNA was prepared, poly(A) RNA was purified, one  
pfu were in vivo excised to give pluescript SK(-) cDNA  
phagemids in the TJ Close lab at the University of  
California, Riverside (Choi). Phagemids were plated and  
picked at the Clemson University Genomics Institute (CUGI)  
(Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA  
preparations, DNA sequencing and sequence analysis were  
performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates  
, Rambo, Main). The sequence has been trimmed to remove  
vector sequence and contains a minimum of 100 bases of  
phred value 20 or above. For more details on library  
preparation and sequence analysis see  
http://www.genome.clemson.edu/projects/barley. To order  
this clone see http://www.genome.clemson.edu/orders Also  
see Close RJ, Wing R, Kleinhoef A, Wise R (2001)  
Genetically and physically anchored EST resources for  
barley genomics. Barley Genetics Newsletter 31:29-30.  
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"  
BASE COUNT 206 a 231 c 207 g 141 t  
ORIGIN

Query Match 5.8%; Score 86.6; DB 12; Length 785;  
Best Local Similarity 53.0%; Pred. No. 9.8e-13;  
Matches 342; Conservative 0; Mismatches 269; Indels 34; Gaps 6;  
QY 633 TCAGGCTCAGGAGATAGGCTGTGTATGAACACTTCGGGTCTCTCCAGCGGATCTCGCA 692  
DB 5 TGTCCGCCCGGAATTCACAGGCCAAGGACACTTCGGATTCAAGGATGGAATCAGTCA 64  
QY 693 GCCTCAGTCACAGCTCGGAGACGCGCTTCCCTGACAGCGGTCTGCCA ---- 747  
DB 65 GCCTCTTCTTGAGGCTGGATGAGAGGTGCGCCAGGCAAGGACCTGTCCGACAAA 124  
QY 748 -CCTGGAATTATCTCACTGAGCGCATGGGGACACGGGACCCGACCGCTCTGGGCTCT 806  
DB 125 GCAGGAATGATGCTTTGCGGATACGACGGCGACGATGCAACAGCCGATGGGCCAA 184  
QY 807 AGATGGAGTTTCATGGCATTTCCGGCACTCCAGCAGAGGTCCCGGAATTCACGCGTA 866  
DB 185 GGACGGCATTTCTTGTCTTCCTGGTATGTCACACAGAGTTCTTGAGTTCGAGAAAT 244  
QY 867 CACGCTCGCCAAACCGATACCCGCGAAGACGCGGGGAAACCTCACCCAGCAGGAAGTGC 926  
DB 245 CATGAACCAAAACGCTAAGCAGCAGCAACATTACCAAGGACACCCCAAGC-----CAGG 298  
QY 927 AGAGTTCTCGGCGCGCATGTTTCGGCGCTTGGAGAGCGCGCGCGCATTTGACCTCGC 986  
DB 299 CGAGAAGCTTCGCGCGTTTACCTGTATGGGTGGTGGAGACCGGAACCCCGCTTGACGATC 358  
QY 987 GCGCAGCGGCGGACGCCAGCGCTCGGCGCGGACCCCGCAGAGGAACAACAAATTTTCATTA 1046  
DB 359 TCCTCAGCATGACACCAAGAGGCTTTGTTGCAATC-----CAACAACTTCGACTT 409  
QY 1047 CTCAGACAGCTTGAGGACGAGACGCGCTGCCCTTCGGTGACACGTGAGGAGACGAA 1106  
DB 410 CGCA---CGGTACCCGAGCAACAAGATGTTCCTTCAGCGCACATCATCCGCAAGATGAG 466

QY 1107 CCCTCGACAGGACCTGGGTGGAGCCGGTCGACACCTTCCAGCTTATGCGGTCCAGTATCCC 1166  
DB 467 ACCCGTGTCTGACTTGG-----AACACGACACCGCGCTCATCTCGTGGGAATCCC 520  
QY 1167 GTACGGCCGAGAAACGCTGATGACAGAACTTGGTCGGCGGTGACTGGCGAA----GACC 1222  
DB 521 CTACGGCAGTGTGTCACTGACGCTGAACTGGCGCGCAGAAAACGCCAAACAAGGAGC 580  
QY 1223 GCGGTCTCTTTTTCGTGAGTACGATCCATTTATTTGGTAAATGGGT 1267  
DB 581 GTGGCCTCTCTCTGTTGTTACCATCATGATTTTCGAAAGGGAT 625  
RESULT 2  
AJ407445 552 bp mRNA linear EST 08-JAN-2002  
LOCUS  
DEFINITION AJ407445 Populus tremula x P. tremuloides/Amanita muscaria mixed  
EST library Populus tremula x P. tremuloides/Amanita muscaria mixed  
EST library cDNA clone 441, mRNA sequence.  
ACCESSION AJ407445  
VERSION AJ407445.1 GI:18091330  
KEYWORDS  
SOURCE Populus tremula x P. tremuloides/Amanita muscaria mixed EST  
library.  
ORGANISM Populus tremula x P. tremuloides/Amanita muscaria mixed EST library  
Eukaryota; mixed EST libraries.  
REFERENCE 1 (bases 1 to 552)  
AUTHORS Nehls,U., Bock,A. and Hampf,Rd.  
TITLE ESTs from Amanita muscaria / Populus tremula x tremuloides  
economycoarhizas  
JOURNAL unpublished (2001)  
COMMENT Contact: Nehls U  
Botanisches Institut/Physiologische Oekologie der Pflanzen  
Universitaet Tuebingen  
Auf der Morgenstelle 1, Tuebingen 72076, Germany.  
FEATURES  
Location/Qualifiers  
Source 1..552  
/organism="Populus tremula x P. tremuloides/Amanita  
muscaria mixed EST library"  
/db\_xref="taxon:143775"  
/clone="441"  
/clone\_lib="Populus tremula x P. tremuloides/Amanita  
muscaria mixed EST library"  
BASE COUNT 124 a 158 c 135 g 132 t 3 others  
ORIGIN  
Query Match 5.8%; Score 83.2; DB 9; Length 552;  
Best Local Similarity 56.2%; Pred. No. 7e-12;  
Matches 239; Conservative 0; Mismatches 174; Indels 12; Gaps 4;  
QY 623 TCAGTGGGTCTCGCGCTCCAGGAGATCAGGCTGGTTCATGAACACTTCGGGTTCCTCGACG 682  
DB 101 TTACTGGAGATGTAAGCCAGGCAAGAGGCTGGACATGAACACTTCGGCTTCTTGGNAC 160  
QY 683 GCA--TCTCGGAGCCCTCAGTCACAGCTGGGAGACGCGTCTTCCCTGGACAGCGGT 740  
DB 161 GGAATCTTCCAAACCCGCGCGCTCATCGGATTCGATACGAATCCCTCTCTGGACCTGTCC 220  
QY 741 CGTCCACCTCGGAATTATCTCTCACTGGACCGCATGGGGACACGG---GCACCCGACCGTC 797  
DB 221 TGTAACGCTCTGGGATCATCTCTCTCGGCCATGATGCGACGATGTATCTACTCGCGATCC 280  
QY 798 GTGGGCTCTAGATGGAGATTTCATGCGATTCGCGCACTTCCAGCAGAGAGTCCCGGAATT 857  
DB 281 CTGGATGTATAGCAAGAAAGTTTCTTGTCTTCCGTTTACCTGTTCACAGCTCGTTCCGAATT 340  
QY 858 CAACCGCTACACGCTCCGCCAACGCGATACCCCGCAACAGCGGGGAAACCTCACCCAGCA 917  
DB 341 CAA--CAAAGTTTCTCCAGGACAAATCCGATCAACGTCGCTGGTCTGTGACAGCCAG-- 396  
QY 918 GGAAGTCAGAGTTTCTCTCGCGCGCGCATGTTCCGCCCTTGGAGAGAGCGGCGCCGAT 977  
DB 397 ---AGGTAGTGAATTCGTCGGTGTCTCGACTTGTATGTCTGGAAGAGCGGTGCTCCAGT 453



OY	978	TGACCTGCAGCCGCGGAGCAGCCAGCGTTCGGCCGACCCGAGAGAACAACA	1037
Db	454	TGAATTGCCACTTTTGACGACGATCAGCCCTTCCGCCCATTCGCCGATAGAACATAA	513
OY	1038	TTTTGG 1042	
Db	514	CTTTGG 518	
<b>RESULT 3</b>			
CN80052P			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
SSS.			
Drosophila melanogaster.			
Drosophila melanogaster			
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
Spheroidea; Drosophilidae; Drosophila.			
1 (bases 1 to 844)			
Genoscope.			
Direct Submission			
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage ;			
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqif@genoscope.cns.fr			
- Web : www.genoscope.cns.fr)			
Determination of this BAC-end sequence was carried out as part of a			
collaboration with the Berkeley Drosophila Genome Project (BDGP).			
The BDGP is constructing a physical map of the Drosophila			
melanogaster genome using these BACS. For further information			
please see http://www.fruitfly.org The BDGP Drosophila			
melanogaster BAC library was prepared by Kazutoyo Osoegawa and			
Aaron Mammeter in Pieter de Jong's laboratory in the Department of			
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,			
NY. The library is named RPCI-98 and was constructed by partial			
Rcort digestion of Drosophila DNA provided by the BDGP from the			
isogenic strain y2; cn bw sp, the same strain used for the BDGP's			
PI and EST libraries. A more detailed description of the library			
and how to order individual BAC clones, the entire library, or			
fillets for hybridization from the BACPAC Resource Center can be			
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.			
Location/Qualifiers			
1..844			
/organism="Drosophila melanogaster"			
/db_xref="taxon:7227"			
/clone_1fb="RPCI-98"			
/note="end : TET3"			
BASE COUNT			
ORIGIN			
261 a 112 c 92 g 35 t 344 others			
Query Match 3.4%; Score 51.4; DB 17; Length 844;			
Best Local Similarity 15.7%; Pred.No. 0.0055;			
Matches 51; Conservative 147; Mismatches 127; Indels 0; Gaps 0.			
OY	763	ACTGAGCCATGTGGGAGCAGCGACCGACCGCATCGTGTGGGCTTAGATGTGAGATTATG	822
Db	391	MCAGCGSSSSASARSRMSSCSCASSCSVSSSSCSAMSSSCCVSVSGMSSSCGS	450
OY	823	GCAATCCGCGCACTTCCAGCAGAAAGTCTCCCGAATTCAAACGGTACAGCTGCCAACGCG	882
Db	451	GMSASSSSSSGGVSVSGGRVAVRGCMCCSNVCCMSKNCMSNVSVACVCSGS	510
OY	883	ATACCCTCGAAGACGCGGGAACCTCACCCAGCAGAAAGGTGACGAGATTCTTGCGCGC	942
Db	511	VRSVAVGCVSGRVGGSRRGARBSRGCGSSVS SVSSSSIVMGWCAACSSASVSGS	570
OY	943	CGCATGTTGGCCGTTGGAAAGACCGCGCGCCGCAATTGACTTCGTGGCGGACGCGACAC	1002

[illegible]

[illegible]

```

QY 461 CCCCAACCTCGGGGACGACCTCAGCCCAATGGTGGCCCTTTTACTGTACCAACCATCC 520
Db 349 CCGATTACCTGATGAAGATCTGACGAGAGCGGGCTACAGCTTACACCAACGCCGAGC 408
QY 521 ATGGTGTCTTTGTGATTGGTACGACGACGACGACTTTTGGATCACTTCCAGATGATA 580
Db 409 GTGAGATCGGGCCGACATCAAGAGAAAGCTGTGTAGTGTGCTTGGATTTGAGAGAG 468
QY 581 TCTCTTCGACCTTTGGTCTTCCATCACTCAAGTGCAGAGCGCTCAGTGGCTCGCCGTC 640
Db 469 AGATGGCCACCGCTCCGCTCCTCAGCTGAGAAATGTGTAGAGTTCTTGAAGAGCC 528
QY 641 CAGAGATCAGAGCTGTGTATGAACACTTCGGATTCTTCCAGAGCATTCGACGCTCAG 700
Db 529 AGGTGATCACCATTGGCAACGAGCGCTTCCGCTGCCGAGCGCTTTCACAGCCCTCG 588
QY 701 TCACAGCTGGGAGACGACCGCTTCTTCCAGAGAGC 737
Db 589 TCTGGGATGAGAGTGTGCGGCACTACGAGACGCTC 625

RESULT 7
LOCUS BE718110 677 bp mRNA linear EST 12-SEP-2000
DEFINITION R1-HT0797-210600-021-g05 HT0797 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE718110
VERSION BE718110.1 GI:10106375
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 677)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,M.Jr., Zago,M.A., Bordin,S., Coستا,F.F.,
Goldman,G.H., Carvalho,A.F., Matukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
COMMENT Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704822
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6&t=R1-HT0797-210
600-021-g05&t3=2000-06-21&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 662.
FEATURES
Source
1..677
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0797"
/dev_stage="Adult"
/note="Organ: head neck; Vector: puc18; Site: 1: Smat;
Site: 2: Smat; A mini-library was made by cloning products
derived from ORFESTS PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research).
Profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 133 a 217 c 191 g 136 t
ORIGIN

```

```

Query Match 3.0%; Score 44.6; DB 12; Length 677;
Best Local Similarity 47.6%; Pred. No. 0.36;
Matches 131; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 461 CCCCAACCTCGGGGACGACCTCAGCCCAATGGTGGCCCTTTTACTGTACCAACCATCC 520
Db 347 CCGATTACCTGATGAAGATCTGACGAGAGCGGGCTACAGCTTACACCAACGCCGAGC 406
QY 521 ATGGTGTCTTTGTGATTGGTACGACGACGACGACTTTTGGATCACTTCCAGATGATA 580
Db 407 GTGAGATCGGGCCGACATCAAGAGAAAGCTGTGTAGTGTGCTTGGATTTGAGAGAG 466
QY 581 TCTCTTCGACCTTTGGTCTTCCATCACTCAAGTGCAGAGCGCTCAGTGGCTCGCCGTC 640
Db 467 AGATGGCCACCGCTCCGCTCCTCAGCTGAGAAATGTGTAGAGTTCTTGAAGAGCC 526
QY 641 CAGAGATCAGAGCTGTGTATGAACACTTCGGATTCTTCCAGAGCATTCGACGCTCAG 700
Db 527 AGGTGATCACCATTGGCAACGAGCGCTTCCGCTGCCGAGCGCTTTCACAGCCCTCG 586
QY 701 TCACAGCTGGGAGACGACCGCTTCTTCCAGAGAGC 735
Db 587 TCTGGGATGAGAGTGTGCGGCACTACGAGACCG 621

RESULT 8
LOCUS CDS0175Y/c 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN37L08 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL108460.1 GI:5628764
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billand at CERH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.
FEATURES
Source
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="BACN37L08"
/clone_11b="DrosBAC"
/plasmid="pBelOBAC11"
/note="end : SP6"
BASE COUNT 254 a 176 c 160 g 152 t 359 others
ORIGIN
Query Match 2.9%; Score 43.4; DB 17; Length 1101;
Best Local Similarity 15.7%; Pred. No. 1;
Matches 70; Conservative 182; Mismatches 195; Indels 0; Gaps 0;

QY 579 TATCTCTTCACCTTTGGTCTTCCATCACTCAGGTGAGCGGCTCAGTGGCTCGCGC 638
Db 1095 TTBSTSTSTSTTTTCTTTTBTBTSSSTSTSTSTBTBSSSSSSBBSGSSSSSSS 1036

```

```
QY 639 TCCAGGATCAGGCTGGTCATGAACACTTCGGGTTCTCCGACGGCATCTCCGACCCCTC 698
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1035 SSSSSSSSSSTSSSSSTSSSSSTSSSSSTSSSSSTSSSSSTSSSSSTSSSSSTSSSS 976
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 699 AGTCACAGCTGGGAGACAGCGCTCTTCCCTGACAGCGGCTGCTCCACCTCGAATAT 758
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 975 TSSSBSTSKSTSSBSTSSSSBSBSSTSSSSBSSTSSSTSSSTSSSTSSSSSSSSSB 916
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 759 CTTCACTGAGCGCATGGGACACGGCGACCCACCGTCTGGTCTCTAGATGGAGTTT 818
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 915 BSTBSBSBTATTTTBTSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSS 856
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 819 CATGCGATTCGGCACTTCAGACAGAGTCCCGGAATTCACAGGTTACAGCTCGCCAA 878
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 855 SYGSSBFTSKSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSS 796
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 879 CGCATATCCCGCAACAGCGCGGAAACCTCACCCAGCAGGAGGTGCAGAGTTCTCTCG 938
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 795 CGKSTSSGTSSTGTCBSCSTGSCSCCCVCCCTCTCTCTCTCTCTCTCTCTCTCT 736
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 939 CGCGGCATGTTCCGCGTGTGGAGAGCGCGCGCGCGATGACCTCGCGCCGACGGCGGA 998
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 735 YSTSSASTGSSCSSTSSSSSGGTGKMGCGCGHTTGTGTCTCTCTCTCTCTCTCTCT 676
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 999 CGACCCAGCGCTCGGCGCCGACCCGCA 1025
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 675 CGSSCCCSYCCGSSSKTSMGTTCM 649
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 9
BE718106
LOCUS BE718106 608 bp mRNA linear EST 12-SEP-2000
DEFINITION RCI-HT0797-210600-021-d04 HT0797 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE718106
VERSION BE718106.1 GI:10106371
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
```

```
REFERENCE
AUTHORS
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Mateukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
```

```
TITLE
JOURNAL
MEDLINE
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml.pl?cl=at2-RC1-HT0797-210
600-021-d04&t3=2000-06-21&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 602.
Location/Qualifiers
1. .658
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0797"
/dev_stage="Adult"
/notes="Organ: head neck; Vector: puc18; Site1: Smai1;
Site2: Smai1; A mini-library was made by cloning products
```

```
FEATURES
source
1. .651
/organism="Homo sapiens"
/db_xref="taxon:9606"
Location/Qualifiers
High quality sequence stop: 565.
Seq primer: puc 18 forward
600-021-d12&t3=2000-06-21&t4=1)
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml.pl?cl=at2-RC1-HT0797-210
600-021-d12&t3=2000-06-21&t4=1)
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml.pl?cl=at2-RC1-HT0797-210
600-021-d12&t3=2000-06-21&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 565.
Location/Qualifiers
1. .651
/organism="Homo sapiens"
/db_xref="taxon:9606"
Location/Qualifiers
```

```
BASE COUNT 118 a 202 c 171 g 117 t
ORIGIN
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
```

```
Query Match 2.9%; Score 43; DB 12; Length 608;
Best Local Similarity 47.6%; Pred. No. 0.95;
Matches 127; Conservative 0; Mismatches 140; Indels 0; Gaps 0;
```

```
QY 461 CGCAAAACCTCGGGAGCAGCTCAGCCAATGGGTGGCGGCTTTTACTGGTACCACCATCC 520
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 333 CCGATTACCTGGAAGATCCTTACGAGCGCGGCTTACAGCTTCACCAACCCCGCAGC 392
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 521 ATGCTGTCTTCTGATTGGTAGCGACGAGGACACTTCTTGATCAGTTACGGATGATA 580
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 393 GTGAGATCGTGGCGACATCAAGAGAAGCTGTGTAGTGGCTCTGGACTTCGAGCAGG 452
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 581 TCTTTTCGACCTTTGGTTCTCTCATCACTCAGTGCAGCGCTCAGTGGTCTGGCGGTC 640
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 453 AGATGGCGACCGCTGCGCGCTCCACCTCGCTGGAGAAGTCGTACGAGTTGCTTACGCGCC 512
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 641 CAGGAGATCAGGCTGTGTCATGAACACTTCGGGTTCCTCGACGGCATCTCGCAGGCCCTCAG 700
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 513 AGGTGATCACCATTGGCAACGAGCGCTTCGCTGCCCGGAGCGCCCTGTTCCAGCCCTCGT 572
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 701 TCACAGCTGGGAGACGACCGCTTTCC 727
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 573 TCCTGGCATGGAGTCGTGGCGCATCC 599
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 10
BE718108/c
LOCUS BE718108 651 bp mRNA linear EST 12-SEP-2000
DEFINITION RCI-HT0797-210600-021-d12 HT0797 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE718108
VERSION BE718108.1 GI:10106373
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
```

```
REFERENCE
AUTHORS
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Mateukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
```

```
TITLE
JOURNAL
MEDLINE
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml.pl?cl=at2-RC1-HT0797-210
600-021-d12&t3=2000-06-21&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 565.
Location/Qualifiers
1. .651
/organism="Homo sapiens"
/db_xref="taxon:9606"
Location/Qualifiers
```

```
FEATURES
source
1. .651
/organism="Homo sapiens"
/db_xref="taxon:9606"
Location/Qualifiers
High quality sequence stop: 565.
Seq primer: puc 18 forward
600-021-d12&t3=2000-06-21&t4=1)
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml.pl?cl=at2-RC1-HT0797-210
600-021-d12&t3=2000-06-21&t4=1)
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml.pl?cl=at2-RC1-HT0797-210
600-021-d12&t3=2000-06-21&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 565.
Location/Qualifiers
1. .651
/organism="Homo sapiens"
/db_xref="taxon:9606"
Location/Qualifiers
```



```
/clone_lib="HT0797"
/dev_stage="Adult"
/notes="Organ: head neck; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 1196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      127 a      210 c      179 g      124 t
ORIGIN

Query Match      2.9%; Score 42.6; DB 12; Length 640;
Best Local Similarity 47.3%; Pred. No. 1.3; Indels 0; Gaps 0;
Matches 129; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 461 CCGCAAACTCTGGGAGACCTCAGCCAAATGGGTGGCGCTTTTACTGGTACCAACATCC 520
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 352 CCGATTACCTCATGAAGATCTGACGGAGCGCGCTACAGCTTCAACCAACCGCGGAGC 411
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 521 ATGGTGTTCTTCTGATTGGTAGCAGCAGACACTTCTTGGATCAGTTTCAGGATGATA 580
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 412 GTGAGATCGTGGCGACATCAAGAGAAAGCTGTGTACGTGGCTCTGGACTTCGAGCAGG 471
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 581 TCTCTTCACCTTTGGTTCTCTCCATCACTCAGGTGCAGGCGCTCAGTGGGTTCGCGTCC 640
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 472 AGATGGCCACCGCTCGCCCTCCACCTCGCTGGAGAAGTCTGTACGAGTTGCTTACGCGCC 531
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 641 CAGGAGATCAGCGTGGTCATGACACTTTCGGTTCTCGACGGCATCTCGACGCCCTCAG 700
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 532 AGTGATCACCATTGGCAACGAGCGCTTCGCTGCCCCGAAGGCCCTGTTTCAGACCTCGT 591
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 701 TCACAGGCTGGGAGACGACCTGTCTTCCTCGGAC 733
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 592 TCCTGGCATGAGTCTGTGGCGCATCCCGAGAC 624
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 13
LOCUS      AM600680      760 bp      mRNA      linear      EST 12-JUN-2002
DEFINITION ESTPMC028 Penaeus monodon's total hemocyte cDNA library (#2)
            Penaeus monodon cDNA clone PMC028 5' similar to actin, mRNA
            sequence.
ACCESSION  AM600680
VERSION    AM600680.1 GI:11037809
KEYWORDS   EST.
SOURCE     black tiger shrimp.
ORGANISM   Penaeus monodon
            Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;
            Penaeidae; Penaeus.
REFERENCE  1 (bases 1 to 760)
AUTHORS   Wongsantichon,J., Boonchoy,C., Udomkit,A., Panyim,S. and
            Sonthayanon,B.
TITLE      Expressed Sequence Tags from Black Tiger Prawn's Hemocytes
JOURNAL    Unpublished (2000)
COMMENT    Contact: Sonthayanon, B.
            Prawn Molecular Biology
            Institute of Molecular Biology and Genetics, Mahidol University,
            Salaya campus
            Phutthamonthon 4 Rd., Phutthamonthon District, Nakhon Pathom, 73170
            , Thailand
            Tel: 66 2 4419003
            Fax: 66 2 4419006
            Email: scbst@mahidol.ac.th
            Seq primer: T3.
            Location/Qualifiers
                1..760
                /organism="Penaeus monodon"
                /db_xref="taxon:6687"
                /clone="PMC028"
                /clone_lib="Penaeus monodon's total hemocyte cDNA library
                (#2)"
```

```
/tissue_type="hemocytes"
/dev_stage="16-17 g average body weight (farm-raised)"
/lab_host="E.coli XL-1-Blue MRF"
/notes="Vector: lambda2AP II; Site 1: EcoR I; Site 2: Xho
I; The library was prepared using protocol given by
supplier, Stragene, Inc."
BASE COUNT      163 a      238 c      182 g      176 t      21 others
ORIGIN

Query Match      2.8%; Score 42; DB 10; Length 760;
Best Local Similarity 47.4%; Pred. No. 2;
Matches 126; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 467 ACCTCGGGAGACCTCAGCAATGGGTGGCGCTTTTACTGGTACCAACATCCATGGTG 526
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 67  ACCTGATGAAGATCTCAGCGAGCGGTGCTACACCTTCAGACCAACCGCGGAGGAGAA 126
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 527 TCTTTCTGATTTGGTAGCAGCAGGACACTTCTTGGATCAGTTTCAGGATGATATCTTT 586
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 127 TCGTGGCGGACATCAAGAGAAAGCTGTGTAGTGGCTCTGGACTTCGAGCAGGAGATGA 186
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 587 CGACTTTGGTTCTCTCATCACTCAGTGCAGGCGCTCAGTGGGTCTGGGGTCCAGGAG 646
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 187 CCACCGCTGCTTCTCTCTCTCGTGGAGAAGTCTTACAGCTCCCTCAGCGCCAGGTGA 246
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 647 ATCAGGCTGTGTCATGAACACTTCGGGTTCCTCGACGGCATCTCGCAGGCCCTCAGTCACAG 706
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 247 TCACCATCGCAACGAAAGGTTCCGGTGGCGGAGGCCCTGTTCCAGGCCCTCATTTCTGTG 306
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 707 GCTGGGAGACGACCGTCTTCCCTGGA 732
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 307 GCATGAGTCTGTGGGCATCCACGAA 332
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
LOCUS      CNS016BR      1201 bp      DNA      linear      GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN15K14 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL106545
VERSION    AL106545.1 GI:56224556
KEYWORDS   GSS.
SOURCE     Drosophila melanogaster.
ORGANISM   Drosophila melanogaster.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 1201)
AUTHORS   Direct Submission
TITLE      Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
JOURNAL
COMMENT    Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk . This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaut at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBelOBAC11.
            Location/Qualifiers
                1..1201
                /organism="Drosophila melanogaster"
                /db_xref="taxon:7227"
                /clone="BACN15K14"
                /clone_lib="DrosBAC"
                /plasmid="pBelOBAC11"
                /notes="end : T7"

BASE COUNT      386 a      182 c      234 g      101 t      298 others
ORIGIN
```



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2002, 14:49:33 (Search time 59 Seconds  
(Without alignments)  
7765.679 Million cell updates/sec

Title: US-09-926-084-8

Sequence: 1 atgcctctgcctgcctgttgc.....tcactgcactatcgtcgtc 1494

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 1533381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing: first 45 summaries

Database: Issued Patents NA.\*  
1: /cgn2\_6/pdata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/pdata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/pdata/1/ina/5A.COMB.seq.\*  
4: /cgn2\_6/pdata/1/ina/5B.COMB.seq.\*  
5: /cgn2\_6/pdata/1/ina/PCTUS.COMB.seq.\*  
6: /cgn2\_6/pdata/1/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43.4	2.9	20235	1 US-07-642-734C-3	Sequence 3, Appli
2	43.4	2.9	20235	3 US-08-439-009A-3	Sequence 3, Appli
3	42.4	2.8	9757	1 US-08-093-453B-1	Sequence 1, Appli
4	42.4	2.8	9759	1 US-08-459-041A-1	Sequence 1, Appli
5	42.4	2.8	9759	3 US-08-999-733-1	Sequence 1, Appli
6	40.2	2.7	13613	4 US-09-105-537-3	Sequence 3, Appli
7	38.8	2.6	2167	2 US-08-461-775-9	Sequence 9, Appli
8	38.8	2.6	2167	2 US-08-461-775-9	Sequence 9, Appli
9	38.8	2.6	2668	2 US-08-461-775-11	Sequence 11, Appli
10	38.8	2.6	2668	3 US-09-031-606-11	Sequence 11, Appli
11	37.2	2.5	538	4 US-08-998-416-1056	Sequence 1056, Ap
12	37	2.5	796	4 US-08-998-416-167	Sequence 467, App
13	36.8	2.5	1248	4 US-09-105-537-7	Sequence 7, Appli
14	36.8	2.5	5970	3 US-09-320-878-21	Sequence 21, Appli
15	36.6	2.4	68750	3 US-09-335-409-1	Sequence 1, Appli
16	36.6	2.4	68750	4 US-09-568-102-1	Sequence 1, Appli
17	36.6	2.4	68750	4 US-09-567-969-1	Sequence 1, Appli
18	36.6	2.4	68750	4 US-09-568-480-1	Sequence 1, Appli
19	36.6	2.4	68750	4 US-09-568-486-1	Sequence 1, Appli
20	36.6	2.4	68750	4 US-09-568-472-1	Sequence 1, Appli
21	36.6	2.4	68750	4 US-09-567-899-1	Sequence 1, Appli
22	35.6	2.4	11220	4 US-09-105-537-32	Sequence 32, Appli
23	35.6	2.4	35081	2 US-08-752-760A-1	Sequence 1, Appli
24	35.6	2.4	36778	4 US-09-105-537-5	Sequence 1, Appli
25	35.6	2.4	38506	3 US-09-320-878-19	Sequence 19, Appli
26	35	2.3	1531	1 US-08-449-986-1	Sequence 1, Appli
27	35	2.3	1531	2 US-08-756-855-1	Sequence 1, Appli

28	34.8	2.3	43280	2 US-08-804-227C-1	Sequence 1, Appli
29	34.6	2.3	2329	4 US-09-411-977-1	Sequence 1, Appli
30	34.4	2.3	1519	1 US-08-225-477B-2	Sequence 2, Appli
31	34.4	2.3	1519	5 PCT-US95-04353-2	Sequence 2, Appli
32	34.4	2.3	11219	1 US-07-642-734C-1	Sequence 1, Appli
33	34.4	2.3	11219	3 US-08-439-009A-1	Sequence 2, Appli
34	34.2	2.3	758	2 US-08-288-630-2	Sequence 1, Appli
35	34	2.3	2900	1 US-08-034-650-9	Sequence 9, Appli
36	34	2.3	2900	1 US-08-449-015-9	Sequence 9, Appli
37	33.8	2.3	666	2 US-08-875-034A-1	Sequence 1, Appli
38	33.8	2.3	2299	4 US-09-153-598A-1	Sequence 1, Appli
39	33.8	2.3	9757	1 US-08-093-453B-1	Sequence 1, Appli
40	33.8	2.3	9759	1 US-08-459-041A-1	Sequence 1, Appli
41	33.8	2.3	9759	3 US-08-999-733-1	Sequence 1, Appli
42	33.6	2.2	30001	1 US-08-125-468-1	Sequence 1, Appli
43	33.6	2.2	30001	2 US-08-474-933-1	Sequence 1, Appli
44	33.6	2.2	53526	3 US-08-658-136-2	Sequence 2, Appli
45	33.6	2.2	53577	3 US-08-658-136-1	Sequence 1, Appli

## ALIGNMENTS

RESULT 1  
US-07-642-734C-3  
Sequence 3, Application US/07642734C  
Patent No. 5824513  
GENERAL INFORMATION:  
APPLICANT: Katz, L  
APPLICANT: Donadio, S  
TITLE OF INVENTION: Recombinant DNA Method for Producing  
TITLE OF INVENTION: Erythromycin Analogs  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Edward H. Gorman  
STREET: Abbott Laboratories D177/APed-2 One Abbott  
STREET: Park Rd  
CITY: Abbott Park  
STATE: IL  
COUNTRY: US  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/642,734C  
FILING DATE: 17-JAN-91  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Danckere, Andreas M  
REGISTRATION NUMBER: 32652  
REFERENCE/DOCKET NUMBER: 4952 US 01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-9396  
TELEFAX: 708-938-2623  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20235 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Saccharopolyspora erythraea  
STRAIN: NRRL 238  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 19..10722

```
OTHER INFORMATION: /codon_start= 19
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19..4470
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 97..1482
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1693..2670
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3406..3921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4171..4428
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..10722
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..5847
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6054..7026
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: dehydratase and enoylreductase domains m"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9433..9984
OTHER INFORMATION: /function= "approximate span
OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
OTHER INFORMATION: 6-deoxyerythronolide B formatio"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc_feature
```

```
LOCATION: 10831..12174
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function= "approximatr span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14062..14610
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14857..15114
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15165..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15172..16569
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16768..17721
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 18379..18921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19149..19398
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19492..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: thioesterase domain of module 6"
US-07-642-734C-3
```

```
Query Match 2.9%; Score 43.4; DB 1; Length 20235;
Best Local Similarity 46.1%; Pred. No. 0.018;
Matches 182; Conservative 0; Mismatches 211; Indels 2; Gaps 1;

QY 845 AGGTCCCGGATTCACGGGTACACGCTGCCAACCGCATCCCGGACAGCGGGGAA 904
DB 12623 AGCCGCGCGCGTCTCGGCCACTCTCGAGGGGAGATCCCGCCGACCTCGCCGCG 12682
QY 905 ACCTCACCACGACGAAAGTGACAGTTCTCGGCGCGCATGTTCGCGCGCTTGGAGA 964
DB 12683 CGCTCACCCTGGAGAGCGCGCCNAAGCTCTCGGTCTCGGAGCGGGTCTGCGCGGC 12742
QY 965 GCGGCGCGCGATTGACTTCGCGCGGACGCGGACGACCGCGTCTCGGCGCGACCGC 1024
DB 12743 TCGGCGCGGACGCGCGCATGTTCGCGGTGGGACCGGACCGAGCGCGCGGACGGA 12802
QY 1025 AGAGGAACAACAATTTTCGATTACTCAGACACGCTGACGACGAGACGCGCTCCCTTCG 1084
DB 12803 TCGGCGCGTTCGCGGCGCGCTCTCCATCGCTCAACGCGCGCGCTCGGTCTCG 12862
QY 1085 GTGCACACGTGAGGAAGACGACCTCGACAGGACCTGGTGAGCGCGTCCACACTT-- 1142
DB 12863 TCGCGGGGAGAGCGGCGCGCTGATGCCGAGTCTATCCCGAGTCCGAGCGCGAAGCATAA 12922
```

```

Db 12932 CGGGGGGGCCGATCCCGTCGACTACAGCCTCCCACTCAACCCAGGTGGATCGGTGGCG 12988
Qy 11443 CCACGCTATCGGCTCCAGATATCCCGTACGCGCCCAAGAAACGCTATGACAACTTGCGTC 1202
Qy 1203 CGGCGTGACTGCGCCAGACGCCGCGGCTTTTTCG 1237
Db 12983 AGGAGCTGCTGACCGAGCTGGCGGGGATCTCCCG 13017
RESULT 2
US-08-439-009A-3
; Sequence 3, Application US/08439009A
; Patent No. 6004787
GENERAL INFORMATION:
APPLICANT: Donadio, S
APPLICANT: Katz, L
APPLICANT: McAlpine, J B
TITLE OF INVENTION: Method of Directing Biosynthesis of
TITLE OF INVENTION: Specific Polypeptides
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSER: Steven F. Weinstein
STREET: Abbott Laboratories D377/AP6D-2 One Abbott
STREET: Park Rd
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/439,009A
FILING DATE: 11-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Casuto, Dianne
REGISTRATION NUMBER: 40,943
REFERENCE/DOCKET NUMBER: 4952.US.D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847-938-3137
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 20235 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722
OTHER INFORMATION: /codon start= 19
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19..4470
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 97..1482
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketacyl ACP synthase of module 3"

```

```

FEATURE:
NAME/KEY: misc_feature
LOCATION: 1693..2670
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3406..3921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4171..4428
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..10722
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..5847
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketocacylACP synthase domain of module"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6054..7026
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: dehydratase and enoylreductase domains m"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9433..9984
OTHER INFORMATION: /function= "approximate span
OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function= "gene =eryA"
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
OTHER INFORMATION: 6-deoxyerythronolide B formattio"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10831..12174
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketocacylACPe synthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function= "approximatr span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14062..14610
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:

```



CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/459,041A  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Greene, Jamie L.  
REGISTRATION NUMBER: 32,467  
REFERENCE/DOCKET NUMBER: 07362-0102  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 818-3700  
TELEFAX: (404) 818-3799  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9759 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-459-041A-1

Query Match 2.8%; Score 42.4; DB 1; Length 9759;  
Best Local Similarity 52.2%; Pred. No. 0.025;  
Matches 94; Conservative 0; Mismatches 86; Indels 0; Gaps 0;  
QY 841 CAGAGGTCCTCCGATTGACGCGGTACACGCTCCGCAACCGATACCCGGAACAGCGCG 900  
DB 5600 CATACGAGACCCGAGGTGATGCTGTGGCGAGGCCCATTAACACCAACGCGCATCGAG 5659  
QY 901 GGAACCTTACCCAGACAGAGGTGACAGTTCTCGCGCGCGCATGTTGCGCCGTTGG 960  
DB 5660 GTGCACTTACTGATTCGATGACATGAAACGACCTCTCTACTGCGGAGCGTGCAGCTCGAG 5719  
QY 961 AAGAGCGGCGCCGCGATTGACCTCGCGCGCGAGCGGAGCAGCAGCGCGCTCGCGCGCGAC 1020  
DB 5720 ATTAGGCGCGCTCTCTTGGGCGCTCCCTTGGCGCGAAGACTACGCGCGCTCGCGCGCGGC 5779

RESULT 5  
US-08-999-733-1  
Sequence 1, Application US/08999733  
Patent No. 6054573  
GENERAL INFORMATION:  
APPLICANT: Frey, Terry K.  
APPLICANT: Pugachev, Konstantin V.  
TITLE OF INVENTION: Highly Infectious Rubella Virus Clones  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew, LLP  
STREET: 191 Peachtree Street, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/999,733  
FILING DATE: 02-SEP-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/459,041  
FILING DATE: 02-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/093,453  
FILING DATE: 19-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/722,334  
FILING DATE: 28-JUN-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Greene, Jamie L.  
REGISTRATION NUMBER: 32,467  
REFERENCE/DOCKET NUMBER: 07362-0103  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 818-3700  
TELEFAX: (404) 818-3799  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9759 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-999-733-1

Query Match 2.8%; Score 42.4; DB 3; Length 9759;  
Best Local Similarity 52.2%; Pred. No. 0.025;  
Matches 94; Conservative 0; Mismatches 86; Indels 0; Gaps 0;  
QY 841 CAGAGGTCCTCCGATTGACGCGGTACACGCTCCGCAACCGATACCCGGAACAGCGCG 900  
DB 5600 CATACGAGACCCGAGGTGATGCTGTGGCGAGGCCCATTAACACCAACGCGCATCGAG 5659  
QY 901 GGAACCTTACCCAGACAGAGGTGACAGTTCTCGCGCGCGCATGTTGCGCCGTTGG 960  
DB 5660 GTGCACTTACTGATTCGATGACATGAAACGACCTCTGCTACTGCGGAGCGTGCAGCTCGAG 5719  
QY 961 AAGAGCGGCGCCGCGATTGACCTCGCGCGCGAGCGGAGCAGCAGCGCGCTCGCGCGCGAC 1020  
DB 5720 ATTAGGCGCGCTCTCTTGGGCGCTCCCTTGGCGCGAAGACTACGCGCGCTCGCGCGCGGC 5779

RESULT 6  
US-09-105-537-3/c  
Sequence 3, Application US/09105537A  
Patent No. 6265202  
GENERAL INFORMATION:  
APPLICANT: Sherman, D. H.  
APPLICANT: Liu, H.  
APPLICANT: Xue, Y.  
APPLICANT: Zhao, L.  
TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
FILE REFERENCE: 600,438US1  
CURRENT APPLICATION NUMBER: US/09/105,537A  
CURRENT FILING DATE: 1998-06-26  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 3  
LENGTH: 13613  
TYPE: DNA  
ORGANISM: Streptomyces venezuelae  
US-09-105-537-3

Query Match 2.7%; Score 40.2; DB 4; Length 13613;  
Best Local Similarity 52.0%; Pred. No. 0.13;  
Matches 90; Conservative 0; Mismatches 83; Indels 0; Gaps 0;



Db 1878 GGCCGCGTCGCCCGATCAAGCCGAGATCGAGTCGACGACTCGGAC 1931

RESULT 9  
US-08-461-775-11

GENERAL INFORMATION:  
APPLICANT: MAZODIER, Philippe  
APPLICANT: GUGLIERI, Gerard  
TITLE OF INVENTION: REGULATORY NUCLEOTIDE SEQUENCE OF THE  
TITLE OF INVENTION: INITIATION OF TRANSCRIPTION  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:

```

? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
?
? CURRENT APPLICATION DATA:
?
? APPLICATION NUMBER: US/08/461,775

```

Query Match	2.6%	Score 38.8;	DB 2;	Length 2668;
Best Local Similarity	47.9%;	Pred. No. 0.16;		
Matches 112; Conservative	0;	Mismatches 122;	Indels 0;	Gaps 0

## RESULT 10

US-09-031-606-11  
; Sequence 11, Application US/09031606  
; Patent No. 6153404

1 APPLICANT: MAZODIER, Philippe  
 2 APPLICANT: GUGLIELMI, Gerard  
 3 TITLE OF INVENTION: REGULATORY NUCLEOTIDE SEQUENCE OF THE  
 4 TITLE OF INVENTION: INITIATION OF TRANSCRIPTION  
 5 NUMBER OF SEQUENCES: 15  
 6 CORRESPONDENCE ADDRESS:  
 7 ADDRESSEE: Burns, Doane, Swecker & Mathis  
 8 STREET: George Mason Bldg., Washington & Prince St.  
 9 CITY: Alexandria

```

:
: MEDIUM TYPE: IBM PC compatible
:
: COMPUTER: IBM PC compatible
:
: OPERATING SYSTEM: PC-DOS/MS-DOS
:
: SOFTWARE: Patentin Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
:
: APPLICATION NUMBER: US/09/031,606
:
: FILING DATE:
:
: CLASSIFICATION:
:
: DEVICE APPLICATION DATA:

```

Query Match	2.6%	Score 38.8	DB 3	Length 2668
Best Local Similarity	47.9%	Pred. No. 0.16		
Matches 112; Conservative	0	Mismatches 122	Indels 0	Gaps 0

RESULT 11  
US-08-998-416-1056  
; Sequence 1056, Application US/08998416  
; Patent No. 6219264  
; GENERAL INFORMATION:  
; APPLICANT: Phillipse, Peter

APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jürgen  
APPLICANT: Knechtle, Philipp  
APPLICANT: Rebschung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESS:  
ADDRESS: No. 6239264artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 1056:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 538 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE: PAG1643UP  
ORGANISM: ASHBYA GOSSYPHII  
US-08-998-416-1056

Query Match 2.5%; Score 37.2; DB 4; Length 538;  
Best Local Similarity 45.5%; Pred. No. 0.24;  
Matches 132; Conservative 0; Mismatches 158; Indels 0; Gaps 0;  
QY 895 ACGCGGAGAACTCACCGACGAGAGGTCGAGAGTTCCTCGCGCGCGCATGTTCCGC 954  
DB 113 ACGCGGAGAGGCTGGCGACGAGCGCGCGCAAAATCGTACTACCGCTGTCGCGAC 172  
QY 955 CATTGGAGAGCGCGCGCGGATTTGACCTCGCGCGGAGCGGACGACCCAGCGCTCGGC 1014  
DB 173 GCGGACAGAGCTCACCGCGCGCACTGCTGCGCGGTGTCCAGCGCGCGGATGACGCGCGCTC 232  
QY 1015 GCCGACCGCGAGAGAAACAATTTTCGATTACTCAGACAGCTGACGAGAGAGCGCGC 1074  
DB 233 TACCGGTACCAACAGCAGATCAGCAAGTCGTTCCAGGACGACCTGATCTACTGCCGCGC 292  
QY 1075 TCGCCCTTCGGTGCACAGTGAAGAGAGAGACGACCTCGAGAGAGAGAGAGAGAGAGAG 1134  
DB 293 GCGCTGCTGAGCAAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 352  
QY 1135 GACACCTTCAGCTATCGGTCGAGTATCCGTCAGGCGCGAGAGAGAGAGAGAGAGAGAG 1184  
DB 353 GAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 402

RESULT 12  
US-08-998-416-467

Sequence 467, Application US/08998416  
Patent No. 6239264  
GENERAL INFORMATION:  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jürgen  
APPLICANT: Knechtle, Philipp  
APPLICANT: Rebschung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESS:  
ADDRESS: No. 6239264artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 467:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 796 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE: PAG1334UP  
ORGANISM: ASHBYA GOSSYPHII  
US-08-998-416-467

Query Match 2.5%; Score 37; DB 4; Length 796;  
Best Local Similarity 49.7%; Pred. No. 0.32;  
Matches 94; Conservative 0; Mismatches 95; Indels 0; Gaps 0;  
QY 861 CCGGTACAGCTCCGCCAACCGGATACCGCGAACAGCGCGGGAACCTTACCCAGCAGGA 920  
DB 262 CGCCAGCGCGCTCGCGCGCTGACCTCTCGCGCGGCTCCGCCCTTGTTCGTGCGCGCTCGC 321  
QY 921 AGGTGACAGAGTTCCTCGCGCGCGCGCATGTTCCGCCCTTGGAAAGAGCGCGCGCGATTGA 980  
DB 322 CGTGCAGCGCTGATCGCGCGCGCTTCCTGTCGCGCGCGACCGCGCGCGCGCTCGC 381  
QY 981 CCTCGCGCGCGAGCGCGGACGACCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1040  
DB 382 CGCGCGCGCGCGCGCGCGCTGTCGCGCGGTAGCGCGCGCGCGCGCGCGCGCGCGCGCG 441  
QY 1041 CGATTACTC 1049  
DB 442 CGTCTGCGC 450

RESULT 13  
US-09-105-537-7



```
; Sequence 7, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600 438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 7
; LENGTH: 1248
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-105-537-7

Query Match      2.5%; Score 36.8; DB 4; Length 1248;
Best Local Similarity 46.2%; Pred. No. 0.45;
Matches 158; Conservative 0; Mismatches 182; Indels 2; Gaps 1;

Qy 727 CCTGCACAGCGGCTGTCTCCACCTGGAATTATCTCTACATGAGCGCATGGAGACACGGGC 786
Db 63 CTTCTGGGGGGGGCCCAACCGCATGACCGGCCCTGTACAGCGCTGACCGGC 122
Qy 787 ACCCGACCGTGTGGGCTCTAGATGGAGATTTCATGGCATTCGCGCATTCACAGAAAG 846
Db 123 CCTGCACAGCGGCTGTCTCCACCGGC--GCCGCTGTCTCCAGATTCTGAGAGCGC 180
Qy 847 GTCCCGAATTCAAGCGGCTACGCTGGCGCAAGCGGATACCGCGAAACGCGGGGAAAC 906
Db 181 GTCCCGGCTGTGGGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGG 240
Qy 907 CTCACCGACAGGAAGGTGACAGATTCTCTGGCGCGGCGCATGTTCCGCCCTTGGAAAGC 966
Db 241 CAGCTCTGTGGCGACCGCGCGGCTCTACCGCGCAATGATGCTGTGATGAGCTTC 300
Qy 967 GCGCGCGGATGACCTCTGGCGCGGCGCGGCGGACGACCGCGGCTGGCGCGGCGGCGG 1026
Db 301 GCGCGCGGCGGCGGCGGCGGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGG 360
Qy 1027 AGGAACACAAATTTTGATTACTCTACAGACACGCTGACGAGCGAG 1068
Db 361 CCGGACACCGGCAACTCTGACCGCGGACGAGGTGGCGCGCG 402

RESULT 14
US-09-320-978-21/c
; Sequence 21, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: Ashley, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
```

```
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 21
; LENGTH: 5970
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-320-878-21

Query Match      2.5%; Score 36.8; DB 3; Length 5970;
Best Local Similarity 46.2%; Pred. No. 0.88;
Matches 158; Conservative 0; Mismatches 182; Indels 2; Gaps 1;

Qy 727 CCTGCACAGCGGCTGTCTCCACCTGGAATTATCTCTACATGAGCGCATGGAGACACGGGC 786
Db 933 CTTCTGGGGGGGGCCCAACCGCATGACCGGCCCTGTACAGCGCTGACCGGC 874
Qy 787 ACCCGACCGTGTGGGCTCTAGATGGAGATTTCATGGCATTCGCGCATTCACAGAAAG 846
Db 873 CCTGCACAGCGGCTGTCTCCACCGGC--GCCGCTGTCTCCAGATTCTGAGAGCGC 816
Qy 847 GTCCCGAATTCAAGCGGCTACGCTGGCGCAAGCGGATACCGCGAAACGCGGGGAAAC 906
Db 815 GTCCCGGCTGTGGGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGG 756
Qy 907 CTCACCGACAGGAAGGTGACAGATTCTCTGGCGCGGCGCATGTTCCGCCCTTGGAAAGC 966
Db 755 CAGCTCTGTGGCGACCGCGCGGCTCTACCGCGCAATGATGCTGTGATGAGCTTC 696
Qy 967 GCGCGCGGATGACCTCTGGCGCGGCGGCGGACGACCGGCTGGCGCGGCGGCGGCGG 1026
Db 695 GCGCGCACCGGCGGCGGCGGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGG 636
Qy 1027 AGGAACACAAATTTTGATTACTCTACAGACACGCTGACGAGCGAG 1068
Db 635 CCGGACACCGGCAACTCTGACCGCGGACGAGGTGGCGCGCG 594

RESULT 15
US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goetlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-335-409-1

Query Match      2.4%; Score 36.6; DB 3; Length 68750;
Best Local Similarity 51.5%; Pred. No. 2.9;
Matches 84; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

Qy 925 GCAGAGTCTCTGGCGCGGCGATTTGGCCGTTGGAAGACCGCGCGGATTTGACTTC 984
Db 64309 GCGCGACCGCTCTGGCGAAGCTGAGCCTCGGCGCTCTGTACCTCTGGCGGTGCTC 64368
Qy 985 GCGCGGACCGGCGGACCGACCGGCTGTGGCGGCGGACCGGCGGAGAAACAAATTTTCAT 1044
Db 64369 GCTCCGCGCGGCGGACATCGGCGCGCGCGCGCGCGCGCGGCGGCGGCGGCGGCGGAT 64428
```

Oy 1045 TACTCAGACACGCTGACGCGACGACGCGCTGCGCCCTTCGGTG 1087  
Db 64429 GACGAGGCCGACGAGGCCGACGAGGCCGCTCCGCCCTTCGGCG 64471

Search completed: December 4, 2002, 23:18:00  
Job time : 247 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2002, 14:49:33 ; Search time 66 Seconds  
(without alignments)  
8826.659 Million cell updates/sec

Title: US-09-926-084-8  
Perfect score: 1494  
Sequence: 1 atgcgcctgcgcctgtctgtt.....tcactcgcatcgcctgct 1494

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0  
350425 seqs, 194966369 residues

Total number of hits satisfying chosen parameters: 700850

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:  
1: /cgn2\_6/prodata/1/pubppna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/prodata/1/pubppna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/prodata/1/pubppna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/prodata/1/pubppna/US07\_PUBCOMB.seq:\*  
5: /cgn2\_6/prodata/1/pubppna/US07\_PUBCOMB.seq:\*  
6: /cgn2\_6/prodata/1/pubppna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/prodata/1/pubppna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/prodata/1/pubppna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/prodata/1/pubppna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/prodata/1/pubppna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/prodata/1/pubppna/US10\_PUBCOMB.seq:\*  
12: /cgn2\_6/prodata/1/pubppna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/prodata/1/pubppna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/prodata/1/pubppna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41.8	2.8	2433	9	US-09-712-363-137 Sequence 137, App
2	40.2	2.7	13613	9	US-09-860-846-3 Sequence 3, Appl
3	40.2	2.7	13613	10	US-09-861-289-3 Sequence 3, Appl
4	38.2	2.6	1584	9	US-09-712-363-10 Sequence 10, Appl
5	38	2.5	467	10	US-09-917-800A-115 Sequence 115, App
6	38	2.5	515	10	US-09-917-800A-73 Sequence 73, Appl
7	38	2.5	2455	10	US-09-917-800A-1386 Sequence 1386, App
8	37.2	2.5	286	10	US-09-294-093B-5229 Sequence 5229, App
9	36.8	2.5	1248	9	US-09-860-846-7 Sequence 7, Appl
10	36.8	2.5	1248	10	US-09-861-289-7 Sequence 7, Appl
11	36.4	2.4	1366	10	US-09-917-800A-1527 Sequence 1527, App
12	35.6	2.4	2208	10	US-09-815-242-7991 Sequence 7991, App
13	35.6	2.4	4071	12	US-10-044-090-591 Sequence 591, App
14	35.6	2.4	11220	9	US-09-860-846-32 Sequence 32, Appl
15	35.6	2.4	11220	10	US-09-860-846-32 Sequence 32, Appl
16	35.6	2.4	35100	10	US-09-782-378A-26 Sequence 26, Appl
17	35.6	2.4	36778	9	US-09-860-846-5 Sequence 5, Appl
18	35.6	2.4	36778	10	US-09-861-289-5 Sequence 5, Appl
19	35.2	2.4	888	10	US-09-815-242-7873 Sequence 7873, App

20	34.8	2.3	419	10	US-09-960-352-12760 Sequence 12760, A
21	34.8	2.3	1929	10	US-09-761-534A-9 Sequence 9, Appl
22	34.6	2.3	1250	10	US-09-815-242-7834 Sequence 7834, App
23	34.6	2.3	2329	9	US-10-102-704-1 Sequence 1, Appl
24	34.6	2.3	2329	10	US-10-057-951-1 Sequence 1, Appl
25	34.6	2.3	2329	10	US-09-084-491A-1 Sequence 1, Appl
26	34	2.3	348	10	US-09-867-550-649 Sequence 649, App
27	33.8	2.3	2034	10	US-09-815-242-8002 Sequence 8002, App
28	33.6	2.2	88429	9	US-09-976-059-1 Sequence 1, Appl
29	33.4	2.2	1536	10	US-09-815-242-4092 Sequence 4092, App
30	33.4	2.2	2107	9	US-09-981-353-49 Sequence 49, Appl
31	33.4	2.2	2712	10	US-09-748-033-4 Sequence 4, Appl
32	33.2	2.2	255	9	US-09-996-634-31 Sequence 31, Appl
33	33.2	2.2	729	10	US-09-755-143-843 Sequence 843, App
34	33.2	2.2	729	10	US-09-780-669-843 Sequence 843, App
35	33.2	2.2	729	10	US-09-822-827-843 Sequence 843, App
36	33	2.2	421	10	US-09-793-306-149 Sequence 149, App
37	33	2.2	513	10	US-09-864-761-8010 Sequence 8010, App
38	32.8	2.2	593	10	US-09-864-761-15852 Sequence 15852, A
39	32.6	2.2	918	10	US-09-878-574-4338 Sequence 4338, App
40	32.6	2.2	1501	10	US-09-768-894A-1 Sequence 1, Appl
41	32.4	2.2	2636	12	US-10-002-600-95 Sequence 95, Appl
42	32.4	2.2	7353	8	US-08-786-531B-1 Sequence 1, Appl
43	32.2	2.2	1140	9	US-09-860-846-15 Sequence 15, Appl
44	32.2	2.2	1140	10	US-09-861-289-15 Sequence 15, Appl
45	32.2	2.2	1433	10	US-09-925-300-614 Sequence 614, App

## ALIGNMENTS

RESULT 1  
US-09-712-363-137  
Sequence 137, Application US/09712363  
Patent No. US/0020164588A1  
GENERAL INFORMATION:  
APPLICANT: Eisenberg, David  
APPLICANT: Rostetter, Sergio H.  
APPLICANT: Marcotte, Edward M.  
TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
FILE REFERENCE: 07419-032001  
CURRENT FILING DATE: US/09/712,363  
PRIOR FILING DATE: 2000-11-13  
PRIOR APPLICATION NUMBER: PCT/US00/02246  
PRIOR FILING DATE: 2000-01-28  
PRIOR APPLICATION NUMBER: 60/179,531  
PRIOR FILING DATE: 2000-02-01  
PRIOR APPLICATION NUMBER: 60/117,844  
PRIOR FILING DATE: 1999-01-29  
PRIOR APPLICATION NUMBER: 60/118,206,  
PRIOR FILING DATE: 1999-02-01  
PRIOR APPLICATION NUMBER: 60/126,593  
PRIOR FILING DATE: 1999-03-26  
PRIOR APPLICATION NUMBER: 60/134,093  
PRIOR FILING DATE: 1999-05-14  
PRIOR APPLICATION NUMBER: 60/134,092  
PRIOR FILING DATE: 1999-05-14  
PRIOR APPLICATION NUMBER: 60/165,124  
PRIOR FILING DATE: 1999-11-12  
PRIOR APPLICATION NUMBER: 60/165,086  
PRIOR FILING DATE: 1999-11-12  
NUMBER OF SEQ ID NOS: 292  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO: 137  
LENGTH: 2433  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
US-09-712-363-137  
Query Match 2.8% Score 41.8; DB 9; Length 2433;  
Best Local Similarity 48.2% Pred. No. 0 005;  
Matches 118; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

```

; 835 TTCCAGCAGAAAGTCCCGCAATTCAACGCGGTACACGCTCGCCCAACCGGATACCCGGAAC 894
; 1711 TCGACACAGGTGGTCCCGCAGGGCTGGCGAAACACCTTCGCAACCGATGACGAAGAC 1770
; 895 AGCGCGGAACCTCACCCAGCAGGAGGTGCGAGAGTTCTCGCGCGCGCGCATGTTCCGC 954
; 1771 GCGGTGGCAGCGGACCGCGCGCGGTTCGCGCGCGCGCGCGGTGGGATCTGCGCGATG 1830
; 955 CGTTGGAAGAGCGCGCGCGCGGATTTGACCTCGCGCGGACGCGGACGACCGCGCTCGGC 1014
; 1831 TCCGCAAAACCGGACACACCGAGCGCACCGGTGCGCGGCTTCGTGGGCTTCACCAAC 1890
; 1015 GCCGACCGCGAGAGAAACAATTTGATTTACTAGACACAGCTGACGACGAGACGGCG 1074
; 1891 CGCTACGCGGGCGGAGAACTACATCTACGACGACTCCAGTTCGCGCGAGATCTGTGTTC 1950
; 1075 TGCCC 1079
; 1951 GGCCC 1955

RESULT 2
US-09-860-846-3/c
; Sequence 3, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-3

Query Match 2.7%; Score 40.2; DB 9; Length 13613;
Best Local Similarity 52.0%; Pred. No. 0.043;
Matches 90; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

; 971 CGCCGATTGACCTCGCGCGGACGCGGACCGGCTCGCGCGCGGACCGGCTCGCGCGGACG 1030
; 11725 CGCGAACTGACCGCGCGGACCGGACCGGCTCGCGCGCGGACCGGCTCGCGCGGACG 11666
; 1031 ACAACAATTTGATTTACTAGACACAGCTGACGACGAGCGGCTCGCGCGGACCGGCT 1090
; 11665 TCCAGCGCCAGATTGGACGCTCCCGATGACCGCGCGGCGGCGGCTTCGCGCGGACG 11606
; 1091 ACGTGAGGAAGACGAACCTTCGACGAGGACCTTGGGTGGACCGGTCGACACCTTC 1143
; 11605 CCGAACGCTGCGCGCGGACCGGAGCGGATCTGGGGCGGCGGCTCCACCGCGTC 11553

RESULT 4
US-09-712-363-10
; Sequence 10, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1584
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-10

Query Match 2.6%; Score 38.2; DB 9; Length 1584;
Best Local Similarity 51.5%; Pred. No. 0.055;
Matches 88; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

; 971 CGCCGATTGACCTCGCGCGGACGCGGACCGGCTCGCGCGCGGACCGGCTCGCGCGGACG 1030
; 11725 CGCGAACTGACCGCGCGGACCGGACCGGCTCGCGCGCGGACCGGCTCGCGCGGACG 11666
; 1031 ACAACAATTTGATTTACTAGACACAGCTGACGACGAGCGGCTCGCGCGGACCGGCT 1090
; 11665 TCCAGCGCCAGATTGGACGCTCCCGATGACCGCGCGGCGGCGGCTTCGCGCGGACG 11606
; 1091 ACGTGAGGAAGACGAACCTTCGACGAGGACCTTGGGTGGACCGGTCGACACCTTC 1143
; 11605 CCGAACGCTGCGCGCGGACCGGAGCGGATCTGGGGCGGCGGCTCCACCGCGTC 11553

RESULT 3
US-09-861-289-3/c
; Sequence 3, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
```

```

; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-3

Query Match 2.7%; Score 40.2; DB 10; Length 13613;
Best Local Similarity 52.0%; Pred. No. 0.043;
Matches 90; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

; 971 CGCCGATTGACCTCGCGCGGACGCGGACCGGCTCGCGCGGACCGGCTCGCGCGGACG 1030
; 11725 CGCGAACTGACCGCGCGGACCGGACCGGCTCGCGCGCGGACCGGCTCGCGCGGACG 11666
; 1031 ACAACAATTTGATTTACTAGACACAGCTGACGACGAGCGGCTCGCGCGGACCGGCT 1090
; 11665 TCCAGCGCCAGATTGGACGCTCCCGATGACCGCGCGGCGGCGGCTTCGCGCGGACG 11606
; 1091 ACGTGAGGAAGACGAACCTTCGACGAGGACCTTGGGTGGACCGGTCGACACCTTC 1143
; 11605 CCGAACGCTGCGCGCGGACCGGAGCGGATCTGGGGCGGCGGCTCCACCGCGTC 11553

RESULT 4
US-09-712-363-10
; Sequence 10, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1584
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-10

Query Match 2.6%; Score 38.2; DB 9; Length 1584;
Best Local Similarity 51.5%; Pred. No. 0.055;
Matches 88; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

; 971 CGCCGATTGACCTCGCGCGGACGCGGACCGGCTCGCGCGCGGACCGGCTCGCGCGGACG 1030
; 11725 CGCGAACTGACCGCGCGGACCGGACCGGCTCGCGCGCGGACCGGCTCGCGCGGACG 11666
; 1031 ACAACAATTTGATTTACTAGACACAGCTGACGACGAGCGGCTCGCGCGGACCGGCT 1090
; 11665 TCCAGCGCCAGATTGGACGCTCCCGATGACCGCGCGGCGGCGGCTTCGCGCGGACG 11606
; 1091 ACGTGAGGAAGACGAACCTTCGACGAGGACCTTGGGTGGACCGGTCGACACCTTC 1143
; 11605 CCGAACGCTGCGCGCGGACCGGAGCGGATCTGGGGCGGCGGCTCCACCGCGTC 11553

RESULT 3
US-09-861-289-3/c
; Sequence 3, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
```

Db 709 GGTACCCCGACGAGGTTACCCCGACGAGGCGGCTACCCCGACGAGGCGGCTAC 768  
Qy 906 CCTACCCAGCAGAGGTTGAGAGTTCTCGGCGCCGATGTTGGCCGTTGAGAG 965  
Db 769 CAGGACCGAGGTCGAGGTTACCCCGACGAGGCGGCGGCTATCGCGCCCTACGAG 828  
Qy 966 CGGCGCGCGGATTCGCTCGCGCGAGCGGACGACCCGAGCGCTGGCGC 1016  
Db 829 CAACGCGCTCTGTTCTCCGCGCGCGGCTCGCGGCTACGCGCTCCCGCG 879

RESULT 5  
US-09-917-800A-115  
; Sequence 115, Application US/09917800A  
; Patent No. US20020119462A1  
; GENERAL INFORMATION:  
; APPLICANT: Mendrick, Donna  
; APPLICANT: Porter, Mark  
; APPLICANT: Johnson, Kory  
; APPLICANT: Castle, Arthur  
; APPLICANT: Elashoff, Michael  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Molecular Toxicology Modeling  
; FILE REFERENCE: 44921-5038-US  
; CURRENT APPLICATION NUMBER: US/09/917,800A  
; PRIOR FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: US 60/222,040  
; PRIOR FILING DATE: 2000-07-31  
; PRIOR APPLICATION NUMBER: US 60/222,880  
; PRIOR FILING DATE: 2000-11-02  
; PRIOR APPLICATION NUMBER: US 60/290,029  
; PRIOR FILING DATE: 2001-05-11  
; PRIOR APPLICATION NUMBER: US 60/290,645  
; PRIOR FILING DATE: 2001-05-15  
; PRIOR APPLICATION NUMBER: US 60/292,336  
; PRIOR FILING DATE: 2001-05-22  
; PRIOR APPLICATION NUMBER: US 60/295,798  
; PRIOR FILING DATE: 2001-06-06  
; PRIOR APPLICATION NUMBER: US 60/297,457  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/298,884  
; PRIOR FILING DATE: 2001-06-19  
; PRIOR APPLICATION NUMBER: US 60/303,459  
; PRIOR FILING DATE: 2001-07-09  
; NUMBER OF SEQ ID NOS: 1740  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO: 115  
; LENGTH: 467  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 AA848563  
US-09-917-800A-115  
Query Match 2.5%; Score 38; DB 10; Length 467;  
Best Local Similarity 57.6%; Pred. No. 0.032;  
Matches 68; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 966 CGGCGCGCGGATTCGCTCGCGCGAGCGGACGACCCGAGCGCTGGCGC 1025  
Db 336 CGCGCGCAAGAACGAGTGGCGCTGAACCCGACGACGCGTGTTCGACCGGAAGCGCT 395  
Qy 1026 GAGGAACAACAATTGCTACTGACAGACGCTGACGAGAGCGGCGGCCCTTC 1083  
Db 396 GATCGCGCCGCAAGTTGCGGACCGCGTGTGCTGAGTGGACATGAGCACTGGCCCTTC 453

RESULT 6  
US-09-917-800A-73/c  
; Sequence 73, Application US/09917800A  
; Patent No. US20020119462A1  
; GENERAL INFORMATION:

; APPLICANT: Mendrick, Donna  
; APPLICANT: Porter, Mark  
; APPLICANT: Johnson, Kory  
; APPLICANT: Castle, Arthur  
; APPLICANT: Elashoff, Michael  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Molecular Toxicology Modeling  
; FILE REFERENCE: 44921-5038-US  
; CURRENT APPLICATION NUMBER: US/09/917,800A  
; PRIOR FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: US 60/222,040  
; PRIOR FILING DATE: 2000-07-31  
; PRIOR APPLICATION NUMBER: US 60/222,880  
; PRIOR FILING DATE: 2000-11-02  
; PRIOR APPLICATION NUMBER: US 60/290,029  
; PRIOR FILING DATE: 2001-05-11  
; PRIOR APPLICATION NUMBER: US 60/290,645  
; PRIOR FILING DATE: 2001-05-15  
; PRIOR APPLICATION NUMBER: US 60/292,336  
; PRIOR FILING DATE: 2001-05-22  
; PRIOR APPLICATION NUMBER: US 60/295,798  
; PRIOR FILING DATE: 2001-06-06  
; PRIOR APPLICATION NUMBER: US 60/297,457  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/298,884  
; PRIOR FILING DATE: 2001-06-19  
; PRIOR APPLICATION NUMBER: US 60/303,459  
; PRIOR FILING DATE: 2001-07-09  
; NUMBER OF SEQ ID NOS: 1740  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO: 73  
; LENGTH: 515  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 AA818604  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(515)  
; OTHER INFORMATION: n = a or c or g or t  
US-09-917-800A-73  
Query Match 2.5%; Score 38; DB 10; Length 515;  
Best Local Similarity 57.6%; Pred. No. 0.034;  
Matches 68; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 966 CGGCGCGCGGATTCGCTCGCGCGAGCGGACGACCCGAGCGCTGGCGC 1025  
Db 380 CGCGCGCAAGAACGAGTGGCGCTGAACCCGACGACGCGTGTTCGACCGGAAGCGCT 321  
Qy 1026 GAGGAACAACAATTGCTACTGACAGACGCTGACGAGAGCGGCGGCCCTTC 1083  
Db 320 GATCGCGCCGCAAGTTGCGGACCGCGTGTGCTGAGTGGACATGAGCACTGGCCCTTC 263

RESULT 7  
US-09-917-800A-1386  
; Sequence 1386, Application US/09917800A  
; Patent No. US20020119462A1  
; GENERAL INFORMATION:  
; APPLICANT: Mendrick, Donna  
; APPLICANT: Porter, Mark  
; APPLICANT: Johnson, Kory  
; APPLICANT: Castle, Arthur  
; APPLICANT: Elashoff, Michael  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Molecular Toxicology Modeling  
; FILE REFERENCE: 44921-5038-US  
; CURRENT APPLICATION NUMBER: US/09/917,800A  
; PRIOR FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: US 60/222,040  
; PRIOR FILING DATE: 2000-07-31  
; PRIOR APPLICATION NUMBER: US 60/222,880  
; PRIOR FILING DATE: 2000-11-02  
; PRIOR APPLICATION NUMBER: US 60/290,029  
; PRIOR FILING DATE: 2001-05-11  
; PRIOR APPLICATION NUMBER: US 60/290,645  
; PRIOR FILING DATE: 2001-05-15  
; PRIOR APPLICATION NUMBER: US 60/292,336  
; PRIOR FILING DATE: 2001-05-22  
; PRIOR APPLICATION NUMBER: US 60/295,798  
; PRIOR FILING DATE: 2001-06-06  
; PRIOR APPLICATION NUMBER: US 60/297,457  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/298,884  
; PRIOR FILING DATE: 2001-06-19  
; PRIOR APPLICATION NUMBER: US 60/303,459  
; PRIOR FILING DATE: 2001-07-09  
; NUMBER OF SEQ ID NOS: 1740  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO: 1386  
; LENGTH: 515  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 AA818604  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(515)  
; OTHER INFORMATION: n = a or c or g or t  
US-09-917-800A-1386



```

; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 1248
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-861-289-7

Query Match      2.5%; Score 36.8; DB 10; Length 1248;
Best Local Similarity 46.2%; Pred. No. 0.14;
Matches 158; Conservative 0; Mismatches 182; Indels 2; Gaps 1;

QY 727 CCTGACAGCGCGCTGCTCCCACTGGAATTATCTCACTGAGCGGATGAGGAGACAGCGGC 786
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 63 CCTGAGGGGCGGCCCAACCGCATCGACCGCGCGCGCTGTACGAGCGGCTCGACCGGGC 122

QY 787 ACCCGACCGTGGGCTCTAGATGGAGTTTCATGCGATTCCGGCACTTCAGCAGAG 846
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 123 CTTGACAGCAGCTGCTGTCTCCAGGCG--GCCGCTGCTCGGAGATTTCAGAGAGCGC 180

QY 847 GTTCCGGAATTACAGCGCTACAGCTCGCCAAAGCGATACCGCGAAACAGCGGGGAAAC 906
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 GTTCCGCGGCTCGCGGGGCTCGGATGCTCGGCTCCACTGCAACGCAAGCGCGGCGCTC 240

QY 907 CTCACCGACGAGAGGTGACAGATCTCTCGCGCGCGGCAATGTTCCGCCCTTGGAGAGC 966
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 CAGCTCTTGGCGCAGCGCGCGGCTTACCGGCAAGTATCATCGCTGATGATGAGCTTC 300

QY 967 GCGCGCGCGATTGACCTCGCGCGGCGAGCGGAGCAGCAGCGCTGCGCGCGAGCCGCGCAG 1026
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 GCGCGCACCCCGCAGCAGCTGGGTGTGATCGGCTTACCGCGGCTTTCGCGCATCGAC 360

QY 1027 AGGAACAACATTTGATTACTCAGACAGCGCTGACGAGCAG 1068
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 361 CCGGACACCGGCACTCGACCCGAGCAGGAGTGGCGCGCG 402

RESULT 11
US-09-917-800A-1527
; Sequence 1527, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Casle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
```

```

; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1527
; LENGTH: 1366
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 Z27118
; US-09-917-800A-1527

Query Match      2.4%; Score 36.4; DB 10; Length 1366;
Best Local Similarity 56.8%; Pred. No. 0.19;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 966 CGCGCGCGGATTGACCTCGCGCGGCGGAGACCCAGCGCTGGCGCGGACCCGCA 1025
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 64 CGCGCGCAAGAACCGGTGGGCTGAACCCGCAACACCGCTTGGAGCGAGCGGCT 123

QY 1026 GAGAAACAACATTTGATTACTCAGACAGCGTACGAGCAGAGACCGCGCTGCTTC 1083
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 124 GATCGCGCGCAAGTTCCGCGAGCCGCTGTGCTGCACTCGGACATGACACTGCGCTTC 181

RESULT 12
US-09-815-242-7991
; Sequence 7991, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7991
; LENGTH: 2208
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2208)
; US-09-815-242-7991

Query Match      2.4%; Score 35.6; DB 10; Length 2208;
Best Local Similarity 45.4%; Pred. No. 0.45;
Matches 128; Conservative 0; Mismatches 154; Indels 0; Gaps 0;
```





Oy 888 CGGAAACAGCGGGGAAACCTCACCCAGACAGAAAGTGCAGAGTTCTCGGCGCGGCAT 947  
| | | | |  
Db 2892 GGTGCTCGCCCGCGCTGTCTCCGCGCGCGCCGACCCGTAAGCTGATGTCCTCCGCT 2951  
Oy 948 GTTCGAGCGTTGGAAAGCGCGCGCGGATTGACCTCGCGCGCGAGAGACCCAGC 1007  
| | | | |  
Db 2952 GGGCGACCGGACGCGCTCGCGGACGCTGGGAGAGCCCTGCGCGCGCGCGGTGAGC 3011  
Oy 1008 GCTCGGCGCGGACC 1021  
| | | | |  
Db 3012 CGTCGACGCGCTCC 3025

Search completed: December 4, 2002, 23:14:05  
Job time : 99 secs

**THIS PAGE BLANK (USPTO)**